PB81-187239

State-of-the-Art Report on Structure-Activity Methods Development

(U.S.) Environmental Research Lab.
Duluth, MN

Apr 81

U.S. Department of Commerce
National Technical Information Service
NTIS.

United States
Environmental Protection
Agency

EPA 600/3-81-029
April 1981

PB81-187239

EPA

Research and Development

STATE-OF-THE-ART REPORT ON STRUCTURE-ACTIVITY METHODS DEVELOPMENT

Prepared for

OFFICE OF TOXIC SUBSTANCES

Prepared by

Environmental Research
Laboratory
Duluth MN 55804

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA 22161

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO.	2.	3. RECIPIENT'S ACCESSION NO.
EPA-600/3-81-029	Problem-Oriented Report	PB81 187239

4. TITLE AND SUBTITLE	5. REPORT DATE
State-of-the-Art of Structure Activity Methods Development	April 1981 6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	**8. PERFORMING ORGANIZATION REPORT NO.**
Gilman D. Veith	
9. PERFORMING ORGANIZATION NAME AND ADDRESS	**10. PROGRAM ELEMENT NO.**
U. S. Environmental Protection Agency Environmental Research Laboratory-Duluth 6201 Congdon Boulevard Duluth, Minnesota 55804	**11. CONTRACT/GRANT NO.**
12. SPONSORING AGENCY NAME AND ADDRESS	**13. TYPE OF REPORT AND PERIOD COVERED**
U. S. Environmental Protection Agency Environmental Research Laboratory-Duluth 6201 Congdon Boulevard Duluth, Minnesota 55804	**14. SPONSORING AGENCY CODE**

15. SUPPLEMENTARY NOTES

16. ABSTRACT The overall objective of this project is to provide the Agency with the technical basis for estimating the toxicity and environmental behavior of organic chemicals from molecular structure. The project is directed toward the evaluation of quantitative structure-activity relationships (QSAR) for use by EPA Program Offices and toward the development of new data and QSAR methods to extend the technique to meet Agency needs.

Specifically, the objective of the Structure-Activity Project is to develop methods to predict the toxicity, persisitence, and treatability of large numbers of untested chemicals using QSAR based on structural, chemical, and biological properties of representative reference data bases. Development of QSAR is being tailored for use in the (1) initial screening of chemicals under TSCA, (2) development of risk assessment strategies, (3) prioritization of chemicals for Water Quality Criteria development, and (4) the optimization of national monitoring programs for toxic chemicals.

This report summarizes the progress during the first six months of the project. The report provides a literature review and perspective for applying structure-activity methods to aquatic toxicity of industrial chemicals. Experimental work centered on developing methods for estimating molecular descriptors such as log P and connectivity indexes and on the development of a systematic structure-activity data base for aquatic toxicity. A new program for entering structures into a computer and calculating connectivity indexes is discussed. A general model for predicting 96 hour LC50 for narcotic chemicals is presented.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS *(This Report)*	21. NO. OF PAGES
Release to Public	Unclassified	116
	20. SECURITY CLASS *(This page)*	22. PRICE
	Unclassified	

EPA 600/3-81-029
April 1981

STRUCTURE-ACTIVITY RESEARCH AT THE ENVIRONMENTAL RESEARCH LABORATORY-DULUTH

State-of-the-Art Report on Structure-Activity Methods Development

October, 1980

Prepared by

Gilman D. Veith

Environmental Research Laboratory-Duluth

6201 Congdon Boulevard

Duluth, Minnesota 55804

List of Contributors:

ERL-Duluth	University of Wisconsin-Superior	University of Minnesota-Duluth
Rosemarie Russo	Daniel Call	Vincent Magnuson
David Weininger	Larry Brooke	Donald Harriss
Brad Greenwood	Nasim Ahmad	Wayne Maanum
Gary Phipps	Carolanne Curtis	Marcia Fulton
Gary Holcombe	Anne Lima	
Dave DeFoe	Steven Lozano	
	Joseph Richter	
	Karen Studders	

Table of Contents

Section	Title	Page
I.	Project Objectives	1
II.	Project Overview	2
III.	Project Organization	6
IV.	Estimation of Acute Toxicity of Organic Chemicals to Fish	8
V.	Mathematical Methods for QSAR	26
VI.	Molecular Descriptor Generation	36
VII.	QSAR Information System	54
VIII.	Activity Reports	57
IX.	Evaluation of Rapid Screening Tests for QSAR	63
X.	Literature Review on Narcosis	88
XI.	References	101

ERL-D Structure-Activity Research Project

I. PROJECT OBJECTIVES

The overall objective of this project is to provide the Agency with the technical basis for estimating the toxicity and environmental behavior of organic chemicals from molecular structure. The project is directed toward the evaluation of quantitative structure-activity relationships (QSAR) for use by EPA Program Offices and toward the development of new data and QSAR methods to extend the technique to meet Agency needs.

Specifically, the objective of the Structure-Activity Project is to develop methods to predict the toxicity, persistence, and treatability of large numbers of untested chemicals using QSAR based on structural, chemical, and biological properties of representative reference data bases. Development of QSAR is being tailored for use in the (1) initial screening of chemicals under TSCA, (2) development of risk assessment strategies, (3) prioritization of chemicals for Water Quality Criteria development, and (4) the optimization of national monitoring programs for toxic chemicals.

II. PROJECT OVERVIEW

The broad objectives of predicting toxicity, persistence, and treatibility of chemicals require considerable definition at the outset. The use of QSAR carries with it the assumption that a well-defined chemical or biological activity can be related to a combination of free-energy, electronic, and steric properties of the molecule. This assumption dictates that the measured activity (hereafter called endpoints) be well-defined and dependent on the chemical rather than the environment into which the chemical is placed. While aquatic toxicity may be estimated directly by predicting the acute LC50 or chronic MATC from reference data bases, estimations of persistence or treatability are likely to be as dependent on the system studied as on the properties of the chemical. Therefore, QSAR in these areas can only be used in predicting rate constants and equilibria which are needed inputs to process models for persistence and treatability. QSAR can be developed to estimate hydrolysis, adsorption, volatilization, and biodegradability - but the full use of this approach will only be realized in the development of transport and transformation process models such as the EXAMS model development at ERL-Athens.

The overall QSAR development process is outlined in Figure 1. Most often, a data base is compiled for the endpoints of interest (acute toxicity, volatility, etc.) and selected molecular descriptors which are thought to be important controlling factors in the relative activities of the chemicals. The data base is often classified into homologous or semi-homologous series of chemicals to increase the likelihood of common modes of action or reaction mechanisms. Where critical data are lacking, a data generation program is needed to provide the data. The compiled molecular descriptors and endpoints are analyzed by a variety of mathematical techniques such as multiple

QSAR PROCESS

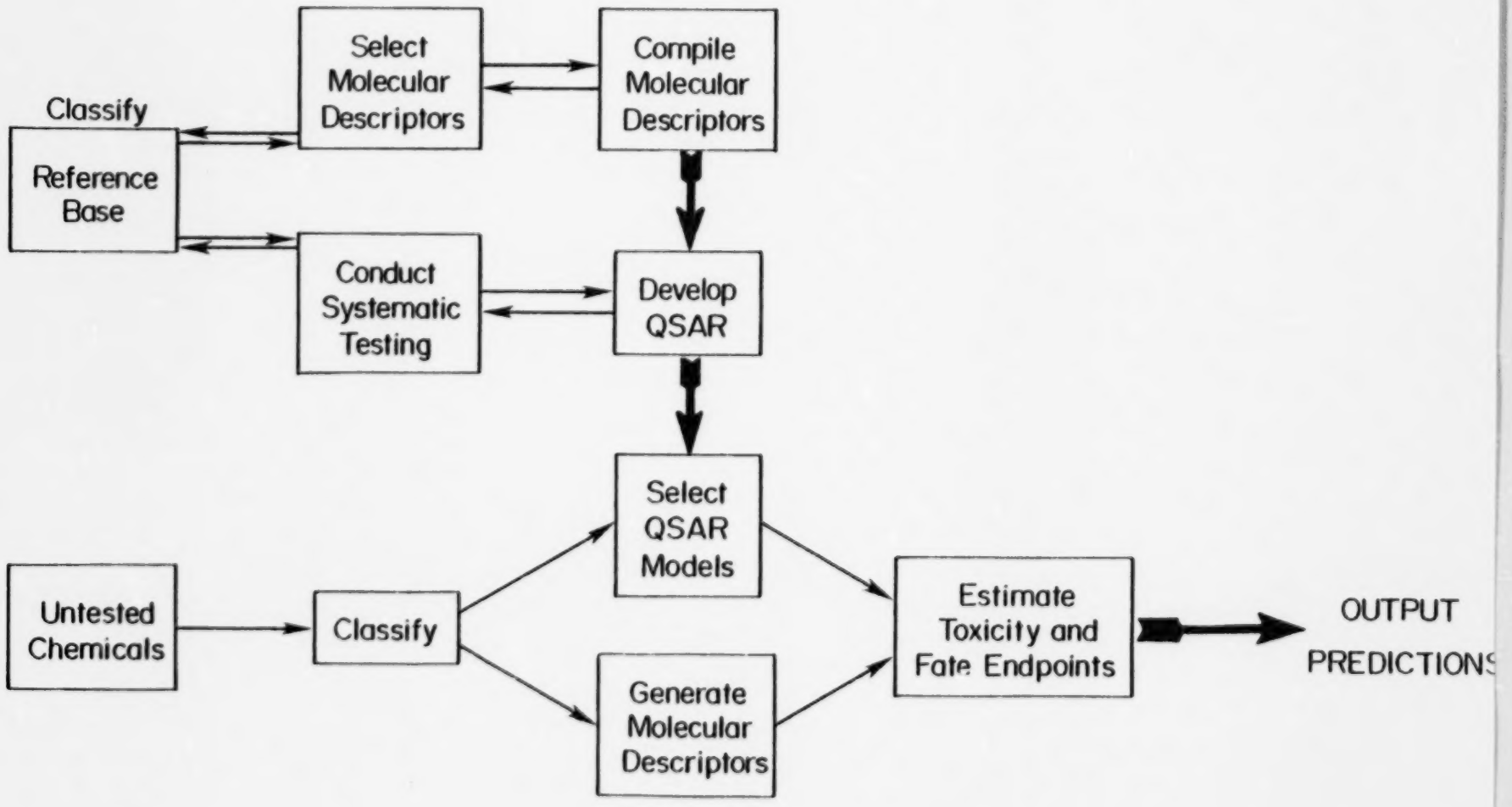


Figure 1. Flow-Chart of the QSAR Process

regression and pattern recognition to derive a quantitative structure-activity relationship for the selected class of chemicals and the specified endpoint.

In making predictions beyond the tested chemicals, the chemical must be classified to select the proper QSAR model. The molecular descriptors which are the variables in the QSAR for the chemical must be generated, and the endpoint is computed as a QSAR prediction. Obviously, only endpoints for which there are QSAR can be predicted.

Figure 1 shows that QSAR must be developed along many research fronts if it is to be useful for large numbers of diverse chemicals. In the initial phase of this project, it was immediately apparent that there were few data of use in QSAR development. The aquatic toxicity literature contained little systematically derived data due to past emphasis on testing relatively few highly hazardous chemicals extensively. Moreover, important molecular descriptors were nearly non-existent for chemicals outside the interests of the drug and pesticide industry. It would be of little value to develop QSAR models which required parameters not available for individual chemicals.

Limitations with respect to the level and mechanism of funding for this project required that the first year objectives be redefined as follows. It was apparent that numerous testing laboratories could develop QSAR models independent of this project if the necessary molecular descriptors were made available to them. Therefore, approximately 25 percent of the resources are assigned to the evaluation, calculation, and compilation of molecular descriptors. An additional 25 percent of the resources are assigned to the development of a user oriented, computerized data dissemination system called the Information System for Hazardous Organics in Water (ISHOW). The

remaining 50 percent of the resources are assigned to the data generation effort.

Because of the expertise of this laboratory in aquatic toxicity testing, data generation is being focused on developing a systematic aquatic toxicity data base for industrial chemicals. In anticipation that the data base will have to contain approximately 20 chemicals of each chemical class, the use of chronic tests is precluded by the cost and time requirements. Therefore, the initial data generation consists primarily of the determination of the 96 hour LC50 of industrial chemicals for fish. Lesser efforts are being made on expanding the model developed to predict the bioaccumulation potential of chemicals in aquatic organisms, and comparing the relative sensitivity of different aquatic species.

It is anticipated that testing will expand to include additional toxicity endpoints after the usefulness of the QSAR technique is demonstrated. Furthermore, coordination of this project with other ORD laboratories which have the expertise in studies of persistence and treatability of organic chemicals is underway. Joint efforts are being discussed for the generation of data bases of key endpoints needed in the modeling efforts in the area commonly called "fate".

III. PROJECT ORGANIZATION

The organization of the ERL-Duluth QSAR project is presented in general terms in Figure 2. The project is centered with a core group of ERL-Duluth staff which is supported by the ORD Decision Unit for Toxic - Ecological Effects and Transport and Fate. The ERL-Duluth staff provides project planning and coordination, tailoring of QSAR methods to meet EPA needs, and a nucleus for toxicity testing. A cooperative agreement between ERL-Duluth and the University of Wisconsin-Superior and funded jointly by ORD and OTS provides the project with data compilation, toxicity data generation, toxicological interpretation of test results and the development of rapid screening methods for the structure activity project. A cooperative agreement with the University of Minnesota-Duluth funded by ORD and EPA-Region V is focused on the dissemination of QSAR predictive technology for large numbers of chemicals, and the development of QSAR models to be tested. The progress of each of the teams listed in Figure 2 will be presented in this report.

Figure 2. Organization of QSAR Project

IV. ESTIMATION OF ACUTE TOXICITY OF ORGANIC CHEMICALS TO FISH

A. Definition of Boundaries

The literature concerning the application of QSAR to environmental sciences is growing rapidly. As with the drug and pharmacology literature, many of the QSAR equations are isolated bits and pieces of information which fail to put any individual study into a larger perspective. This fragmentation becomes especially frustrating when the literature is searched for possible solutions to the problems facing EPA. In considering the problem of estimating the behavior of 30,000-40,000 industrial chemicals in the aquatic environment, it seems necessary to approach the literature with the belief that such things as toxicity and persistence follow very simple relationships among chemicals and that there are some boundary conditions to the problem. Without this belief, our future generations will be continuing our testing in the "hit and miss" fashion of the past 30 years.

In 1968, Corwin Hansch (Hansch, et al. 1968) provided an illustrative example which demonstrated that organic chemicals in water behaved in an orderly fashion. Due to the relationship between factors controlling dissolution of organic chemicals in water and that of partitioning the chemical between a water phase and a lipid phase, Hansch demonstrated quantitative relationships between the water solubility of chemicals and their n-octanol/water partition coefficients. The relationships were simple:

$$\log \frac{1}{s} = a \log P + b$$

where s is the solubility in moles/l, P is the partition coefficient, and a and b are constants. Using data for 156 chemicals of 9 chemical classes, the equation

$$\log \frac{1}{s} = 1.339\ (\pm\ 0.07) \log P - 0.978\ (\pm\ 0.15) \qquad r = 0.935$$

held over a range of about 100,000 in water solubility. When the data were separated into individual classes, values of r were as great as 0.990.

Therefore, it is apparent that log P is a free-energy term which is a manifestation of much of the structural information which affects the behavior of chemicals in aqueous systems.

Implicit in these relationships is the likelihood that the log P will also be important in the partitioning of chemicals from water into aquatic organisms and onto organic-coated surfaces. Therefore, it is almost anticlimactic to have published papers like that of Veith et al. (1978) showing linear relationships between the bioconcentration factor in aquatic organisms and log P, or that of Karickhoff et al. (1978) showing relationships between the sediment/water partition coefficients and log P as follows:

$$K_p = 0.6 \times P \times (O.C.)$$

where K_p is the sediment/water partition coefficient and O.C. is the fractional mass of organic carbon with sediment. Lopez-Avila and Hites (1980) reported that the concentration of numerous organic chemicals in river sediments downstream from discharges can be modeled with the empirically derived equation

$$\log \frac{C_0}{C} = b_0 + b_2 \, (\text{distance}/\log P)$$

where C_0 is the concentration in the discharge and b_0 and b_2 are constants for the river system.

The work presented by Hansch is important to aquatic toxicity studies because it helps to define the boundaries of the "testing space". For example, Figure 3 presents a plot of the "testing space" for chemicals over the log P range from -2 to 8. This is 10 orders of magnitude of lipophillicity and includes the vast majority of industrial organic chemicals on the TSCA inventory. The vertical axis in Figure 3 is the logarithm of the concentration of chemicals expressed as moles/liter which varies from 2 to -8,

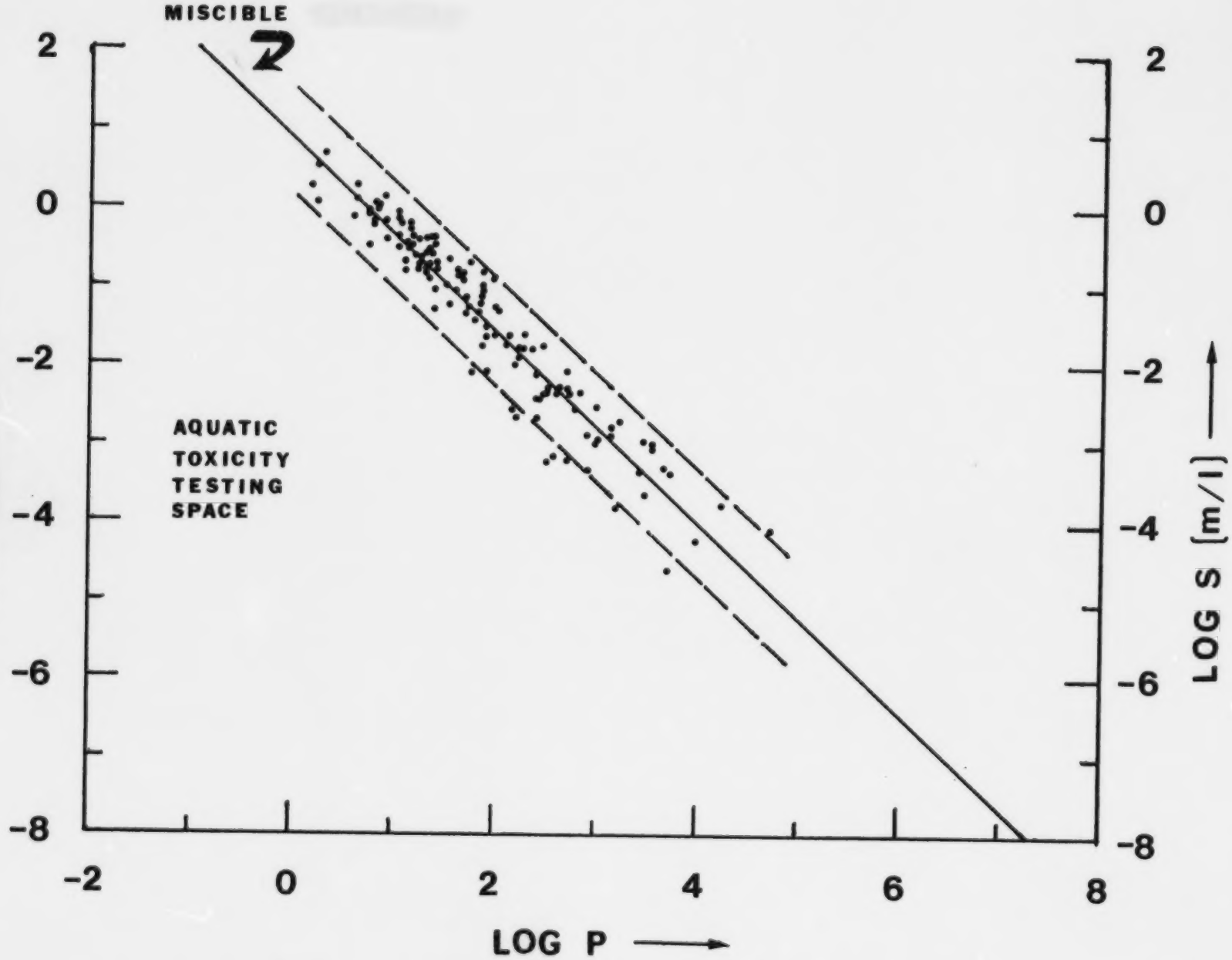


Figure 3. Aquatic Toxicity Testing Space Boundaries

which is approximately the range of concentrations of interest in testing for most toxic effects in the aquatic system. The right-hand scale is the water solubility of chemical expressed as the logarithm of the molar concentration at saturation. The line represents the equation presented by Hansch reanalyzed to include the 95 percent prediction limits of the relationship between water solubility and log P. This "testing space" should enclose most of the organic chemicals encountered in this study. Quite obviously, if the estimated LC50 for a chemical of known log P lies above the water solubility line, it should not be possible to measure the LC50 in aquatic systems.

B. Model for Physical Toxicity in Acute Tests

Little has been done in the area of predicting the toxicity of organic chemicals to aquatic organisms from structure since the work of Meyer and Overton in 1899-1901. The literature contains a few isolated examples demonstrating toxicity-structure relationships; however, a perspective suitable for shaping research to address the problems facing EPA, and OTS in particular, is absent.

There is a thread running through the literature which may give the necessary guidance in tackling the problem. Richet's Rule published in 1893 proposed that the toxicity of a chemical is roughly inversely proportional to its water solubility. Implicit in this rule is the fact that comparable modes of action are considered. Ferguson (1939) proposed that when a physiological effect is reversible, when an equilibrium exists between the organism and the external phase, and when the physiological effect is a function of the external concentration, the toxic effect must be physical in character. He suggested that all substances at some concentration will exert toxic

action by physical mechanisms and even nitrogen gas at high enough pressure can produce effects. The physical effect may be masked by irreversible chemical effects; but in the absence of chemical effects, the primary process is reversible and physical in nature.

The physical effect of Ferguson was expressed as

$$C = kS^{\frac{1}{n}}$$

where S is the water solubility expressed in moles/liter. This physical effect is called narcosis which is any reversible decrease in physiological function induced by physical or chemical agents (Mullin, 1954). Ferguson proposed that the chemical activity for chemicals at concentrations producing narcosis is constant. Since the thermodynamic activity of a chemical producing narcosis (A_{nar}) is calculated by

$$A_{nar} = \frac{C_{nar}}{C_S}$$

where C_{nar} is the concentration producing narcosis and C_S is the solubility of the chemical in water expressed as moles/liter, if the activities of chemicals for the same physiological effect are constant for series of chemicals, the observed effects should be a constant ratio with respect to the water solubility as expressed by Richet's Rule.

Mullin (1954) presented data consistent with this idea and further elaborated on the nature of physical and chemical toxicity. He defined narcosis as any reversible decrease in physiological function where the effects can be divided into narcosis of cell division and narcosis of the central nervous system (general anesthesia). Most chemicals can cause both physical and chemical toxicity; however, if the chemical toxicity is small under the exposure time studied, the predominant effect will be physical

toxicity or narcosis as the exposure concentration approaches the water solubility of the chemical. Thus, in a simplistic way, the prediction of toxicity of chemicals to fish may parallel the zero based budgeting (ZBB) method in that, unless specific functional groups that produce chemical toxicity are present in a chemical of interest, the chemical is likely to cause physical toxicity, or narcosis, at some fraction of its water solubility. If chemical toxicity occurs, the physical toxicity is masked and the observed toxic effects are likely to be at lower concentrations than the concentration needed for narcosis.

The model immediately posed a problem for this project. If the thermodynamic activity at equipotency for narcosis is constant within a series, it would be impossible for higher members of the series (i.e. higher log P) to be non-lethal at water solubility. However, it is a common occurrence in toxicity tests with fish to find chemicals non-lethal at solubility, with hexachlorobenzene and tridecanol being prime examples. Glave and Hansch (1972) provide the answer to this apparent discrepancy. They presented evidence that the linear relationship between narcosis and lipophillicity as measured by log P (inversely proportional to water solubility) was only linear for the more water soluble chemicals. The relationship between toxicity and log P was actually parabolic in nature and was termed the "parabolic effect". The biological effect is actually modeled by:

$$\log\frac{1}{C} = -a(\log P)^2 + b \log P + c$$

Thus, at small values of log P, the equation may appear linear. However, as the range of log P is extended, the relationship between biological activity and log P reverses and the toxicity actually decreases with increasing log P.

Since the log P and water solubility are inversely but linearly related, the expected physical toxicity for chemicals with a large log P could be above the expected water solubility and no effects would be measured.

The "parabolic effect" has been interpreted in many ways. The fact that toxicity decreases with increasing log P at the high end of the log P scale seems to indicate that the chemicals either diffuse slower through membranes or that the fatty tissue act as a storage media and remove the chemical from the blood and key receptors. Since the bioconcentration process is linear with log P, chemicals of high log P may produce large residues in fish at water concentrations which are not toxic to the fish. This is classically illustrated with PCBs where the hazards to consumers of fish due to the accumulated residues far outweight the hazards to fish at environmental concentrations.

The "parabolic effect" may be more accurately described as a bilinear effect resulting from competing membrane transport phenomena. These models have been described by Yalkowsky and Flynn (1973) and Kubinyi (1979). The bilinear model is expressed as:

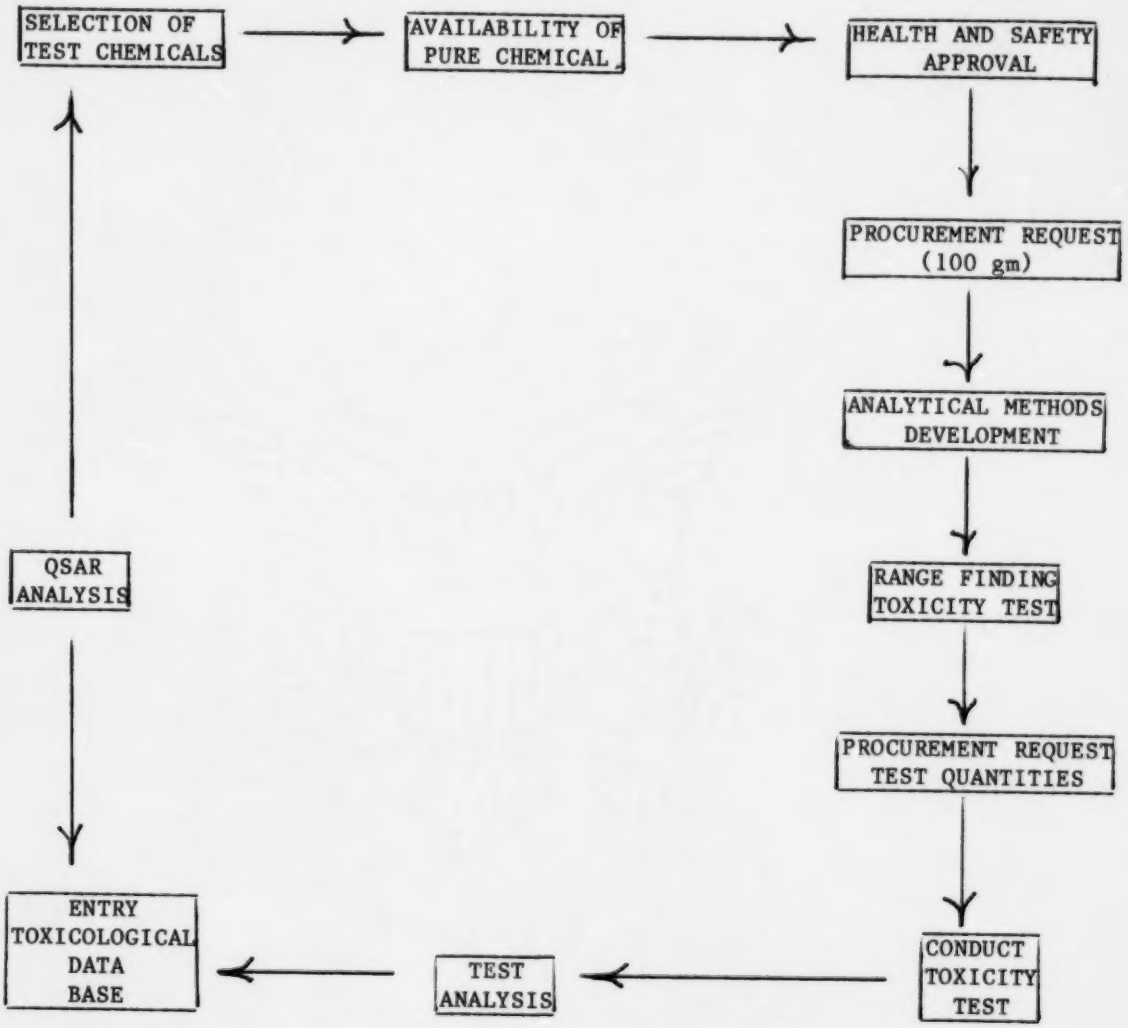
$$\log \frac{1}{C} = a \log P - b \log(\beta P + 1) + c$$

This is more consistent with the understanding of membrane transport which suggests that water-soluble-chemical transport may be controlled by the membrane while lipid-soluble-chemical transport may be controlled by the diffusion layer.

C. Experimental Procedures

The chemicals initially selected for this study include a wide variety of alcohols, ketones, aldehydes, ethers, phenols, and chlorinated aliphatic and aromatic hydrocarbons. Figure 4 illustrates the procedure for selecting

Figure 4. Flow-Chart for Selecting Test Chemicals



chemicals for QSAR modeling. The QSAR modeling group selects appropriate classes of chemicals for consideration. The classes of chemicals are mapped onto a subset of the TSCA inventory to select chemicals which meet the QSAR criteria and are in actual production in industry. The list is further reduced to those which can be purchased as reasonable cost in a pure state. The list is then screened by the ERL-D Health and Safety Officer to insure worker safety. Small quantities of the chemicals are purchased to develop suitable analytical methods and conduct range finding tests so that an appropriate amount of chemical can be purchased. After the test is conducted, evaluated and entered into the data base, the data are used to determine if additional chemicals in the class should be tested to extend the predictive capabilities of the QSAR models. The entire process in Figure 4 requires approximately 4 to 5 months. Therefore, there is not the rapid feedback needed for a cost-effective testing program and this project is optimizing the testing resources by testing several classes of chemicals in a 4 to 5 month "leap-frog" manner. This permits the testing teams to test initial chemicals of one class while the modeling team is selecting additional chemicals of another class.

Exposure Systems. Tests were conducted in proportional diluters (Mount and Brungs, 1967) each with a dilution factor of 0.6. Each diluter delivered 12 L every 20 min cycle to six flow splitting chambers which delivered 0.5 L to duplicate test chambers. Tests chambers were glass aquaria measuring 20 x 35 x 25 cm with a 9 cm standpipe, providing a 6.3 L volume. Fluourescent bulbs provided 28-34 ft candles of light at the water surface. A 16 hr light, 8 hr dark photo period was used.

Water Characteristics. The water source was Lake Superior water maintained at 25°C in the test chambers. Hardness and alkalinity were either measured in the control tanks during the test or from a separate line weekly. pH was measured once during each test at three concentrations and in the control. Dissolved oxygen was also measured at three concentrations and the control at 0, 48, and 96 hours. Ten tests yielded a mean hardness of 56.3 mg/L $CaCO_3$, a mean alkalinity of 42.2 mg/L $CaCO_3$ and an arithmetic mean pH of 7.5. Dissolved oxygen was maintained at 60% saturation or higher. Other chemical characteristics of Lake Superior water are presented in Table 1.

Toxicant Additions. Toxicants were usually introduced directly into the diluter by a metering pump (FMI Corp.). Some tests required delivery of a saturated solution of the toxicant in lake water. The saturator systems consisted of either a 20 L glass tank or 20 L glass jar which contained constant volume of water and excess of the toxicant. The water was stirred continuously to maintain saturation in the incoming water as the saturated solution was pumped to the diluter. Two or three tanks and/or jars, connected with siphons were installed when toxicity was very close to saturation.

Test Organisms. All tests were conducted with fathead minnows (Pimephales promelas) from the Environmental Research Laboratory-Duluth culture units. Fish were hatched and reared in Lake Superior water and fed live brine shrimp at least twice daily. Typical fish tested were 30 days old and weighed 0.12 g. Twenty-five fish were randomly assigned to each of twelve tanks in lots of five. Fish were not fed during the 96 hour tests. Deaths were recorded at 1, 3, 6, 12, and 24 hr and every 24 hrs thereafter

Table 1. Chemical characteristics of dilution water from Lake Superior (all values in µg/liter except where noted otherwise)

pH (pH units)	7.7	Chromium	<0.1
Hardness	46,000	Cobalt	<0.4
Alkalinity	40,000	Copper	1.0
Chloride	1,300	Iron	28.0
Sulfate	3,800	Lead	<0.2
Total phosphate	2	Magnesium	1,400
Ammonia-nitrogen	N.D.[a]	Manganese	5.6
Nitrite-nitrogen	1	Mercury	<0.01
Nitrate-nitrogen	220	Nickel	<0.5
Silica	2,100	Potassium	600
Specific conductance (µmho, 25°C)	93	Silver	<0.02
Aluminum	8	Selenium	0.6
Arsenic	0.7	Sodium	1,300
Cadmium	0.04	Zinc	0.9
Calcium	13,000		

[a] N.D. = not detected

for the remainder of the test. Other observations such as changes in equilibrium, schooling, and distribution of the fish in the tank were recorded.

Data Analysis. Lethal effects of the toxicants were determined by calculating LC50 values at several time intervals for each test. All calculations were made on the ERL-D's PDP-11 computer using the trimmed Spearman-Karber Method (Hamilton et al., 1977).

Analytical Techniques for Toxicity Test Chemicals. Concentrations of toxicity test chemicals are regularly determined in the exposure chambers during each test. Water samples for the polar chemicals were collected and analyzed either by direct aqueous injection on a gas-liquid chromatograph (GLC) or by UV-spectroscopy. GLC analysis was performed on a Hewlett Packard Model 5730A gas chromatograph with dual flame ionization detectors. The instrument was equipped with a 122 cm x 2 mm I.D. column packed with the porous polymer, Tenax-GC (Applied Science, Riviera Beach, Fla.). The flame detector was operated at 300°C and inlet temperatures were either 200 or 250°C depending on the compound being analyzed. Nitrogen was used as the carrier gas and hydrogen and air were used for flame operation. Nitrogen, hydrogen, and air flow rates were 15, 25, and 240 $ml \cdot min^{-1}$, respectively.

Peak area calculations were performed by a Hewlett-Packard Laboratory Automation Data System. Gas chromatographic parameters for each compound analyzed are listed in Table 2.

Standards for alcohols, ketones and aldehydes were prepared in Lake Superior water or distilled water just prior to each test. No differences were noticed in standards prepared in either lake water or distilled water.

Table 2. GLC Parameters and Analytical Quality Control Results for Test Compounds

Test Compound	Isothermal GLC Temp (C)	Retention[1] Time (Min)	Attenuation	Working[2] Range ($\mu g/ml^{-1}$)	Mean % Recovery X ± s.d. n = 4-5	% Agreement of Duplicates n = 4-5
Ethanol	100	0.95	5,120	2,000-25,000	99.3 ± 1.2	98.1 ± 1.4
2-Propanol	100	2.11	1,600	1,800-30,000	-	97.6 ± 2.2
1-Butanol	130	2.80	640	300-10,000	-	97.4 ± 2.2
1-Hexanol	180	1.13	40	15-400	103.6 ± 9.1	95.6 ± 3.6
1-Octanol	185	3.52	3	30-32	94.6 ± 4.9	97.8 ± 2.2
1 Decanol	205	3.9	2	0.6-10	104.6 ± 6.2	97.1 ± 3.3
1-Methyl-1-propanol	130	1.1	160	60-1,900	95.8 ± 4.6	94.8 ± 3.8
2-Chloroethanol	140	1.7	8	12-300	104.1 ± 18.3	95.0 ± 5.3
2,2,2-Trichloroethanol	180	1.6	16	48-480	94.8 ± 9.0	97.3 ± 2.9
2,3-Dibromopropanol	205	2.7	8	5-130	99.4 ± 10.6	94.2 ± 4.3
2-Methyl-2,4-pentanediol	200	1.0	12,800	2,500-25,000	100.3 ± 1.3	98.9 ± 0.9
Cyclohexanol	180	1.7	40	70-1,200	100.7 ± 1.02	98.7 ± 0.4
2-Phenoxyethanol	225	2.1	10	25-900	102.8 ± 3.9	94.8 ± 4.4
Acetone	100	1.95	3,200	1,500-25,000	105.7 ± 2.2	99.0 ± 0.3
2-Butanone	130	1.6	640	600-5,000	110.8 ± 4.1	98.4 ± 1.0
2-Octanone	185	2.1	8	5-80	121.4 ± 4.4	90.0 ± 11.3
2-Decanone	205	2.9	2	0.4-12	89.7 ± 9.6	90.1 ± 8.8
4-Methyl-2-pentanone	165	1.1	160	220-4,400	104.4 ± 3.0	97.2 ± 1.4
5-Methyl-2-hexanone	165	2.0	40	120-2,000	122.4 ± 2.8	92.5 ± 6.6
6-Methyl-hepten-2-one	200	1.54	16	18-200	112.2 ± 5.2	98.8 ± 0.8
2,4-Pentanedione	165	1.8	8	7-390	108.5 ± 13.8	93.3 ± 4.5
Cyclohexanone	180	2.0	80	80-1,300	101.6 ± 2.3	98.6 ± 1.1
Cyclohexanone	180	2.0	80	80-1,300	100.3 ± 3.9	98.0 ± 1.0
Ethanol	80	1.6	4	10-200	104.4 ± 1.7	99.3 ± 0.8
Butanol	120	2.3	2	10-120	99.6 + 3.3	96.9 + 1.9

[1] Retention time measured from point of injection.
[2] Working range is dependent on toxicity exposure levels and not on GLC capabilities.

For water soluble compounds (>100 ppm) standards were prepared by direct addition of the compounds into water. Stock solutions of compounds with lower water solubility (<100 ppm) were made in acetone or methanol. Working standards were made by diluting the stock solutions with water. The solvent concentration in the aqueous working standards was kept below 10% (V/V). Quality control consisted of a duplicate and a spiked water sample with every set of 6 to 12 samples. Spikes were prepared in the same manner as the standards. The spikes provided information on the accuracy of the standards, the stability of standards with time, and the reproducibility of spike preparations.

D. Experimental Results

During the first phase of this project, in excess of 80 chemicals have been tested. The absence of molecular descriptors such as log P in the literature precludes a detailed analysis of all data at this time. However, to test the hypothesis described in the previous sections, a subset of 37 alcohols, ketones, ethers and alkyl halides which appeared to cause death by physical toxicity and for which log P estimates could be made were considered. The 96-hr LC50 for these chemicals are presented in Table 3. The data show that the LC50 (expressed as mg/l) ranged from 28,200 for methanol to 1.53 for hexachloroethane. The log P for the chemicals ranged from approximately -0.85 for methanol to 5.6 for tridecanol. The 96 hr LC50 for tridecanol is greater than the water solubility of this chemical.

The data in Table 3 are plotted (converted to moles/liter) vs. the respective log P values in Figure 5 which is similar to the general description of the aquatic "testing space" described in Figure 3. The water solubility line is calculated from the equations presented by Hansch, excluding alkanes. Within experimental error, the 96 hr LC50 decreases

Table 3. Summary of Acute Toxicity Tests

96 Hour Tests with Fathead Minnows

Chemical Name	LC50 (mg/L)
N-Butyl ether	32.5
Phenyl ether	4.0
Tetrachloroethylene	13.5
1,1,2,2,-Tetrachloroethane	20.3
Pentachloroethane	7.3
1,1,2-Trichloroethane	81.7
1,1,2-Trichloroethylene	44.1
1,2-Dichloropropane	139
1,2-Dichloroethane	118
Hexachloroethane	1.5
1,3-Dichloropropane	131
Methanol	28,200
Ethanol	14,800
2-Propanol	9,640
1-Butanol	1,740
1-Hexanol	97.2
1-Octanol	13.4
2,2,2-Trichloroethanol	298
2,3-Dibromopropanol	71
2-Methyl-1-propanol	1,460
2,4-Pentanedione	96.0
2-Methyl-2,4-pentanediol	10,700

Table 3. (Continued)

Chemical Name	LC50 (mg/L)
Cyclohexanol	705
2-Phenoxyethanol	346
1-Decanol	2.3
Tridecanol	No toxicity at solubility
Acetone	8,140
2-Butanone	3,200
2-Octanone	37.0
5-Methyl-2-Hexanone	158
4-Methyl-2-Pentanone	509
Cyclohexanone	630
2-Decanone	5.7
3,3-Dimethyl-2-butanone	86.0
5-Nonanone	31.0
Acetophenone	161
3-Pentanone	1,530

linearly with increasing log P (i.e. toxicity increases linearly with log P) at least until log P = 3 is reached. In this region of the testing space for these chemicals producing narcosis, the 96 hr LC50 appears to be a constant proportion of the expected water solubility as suggested by the previous discussion. Although, it is premature to propose a constant factor to estimate toxicity of narcotics, the LC50 for the group of chemicals appears to be approximately 0.017 of the water solubility.

The LC50 values for chemicals which have log P values near 4 appear to be much closer to the water solubility. It is proposed that this is the region where the "parabolic effect" is evident. The data point plotted just above the solubility line is that of tridecanol which is non-lethal to fish in a saturated solution of the chemical. The LC50 for decanol is below water solubility (log P = 4.07) which suggests the "toxicity line" and the "water solubilty line" cross between log P of 4 and 5.5. Chemicals with log P values in this range are scheduled for testing next month to better define the toxicity relationship in this range. At the same time, a computer program incorporating the bilinear model is being developed specifically for these data.

In preparing Figure 5, seven chemicals appeared to have anomalous behaviour. Four of these were found to have incorrect log P values reported including an apparent typographical error of log P = 1.70 for decanol. Corrections of these errors removed apparent discrepancies. The apparent outlying point in Figure 5 with an LC50 of -3.0 is that of 2,4-pentanediol. The log P value for this chemical was estimated using the substituent constant approach. Since 2-methyl-2,4-pentanediol is not anomolous, it is reasonable to expect these discrepancies can be evaluated when more reliable log P values are available.

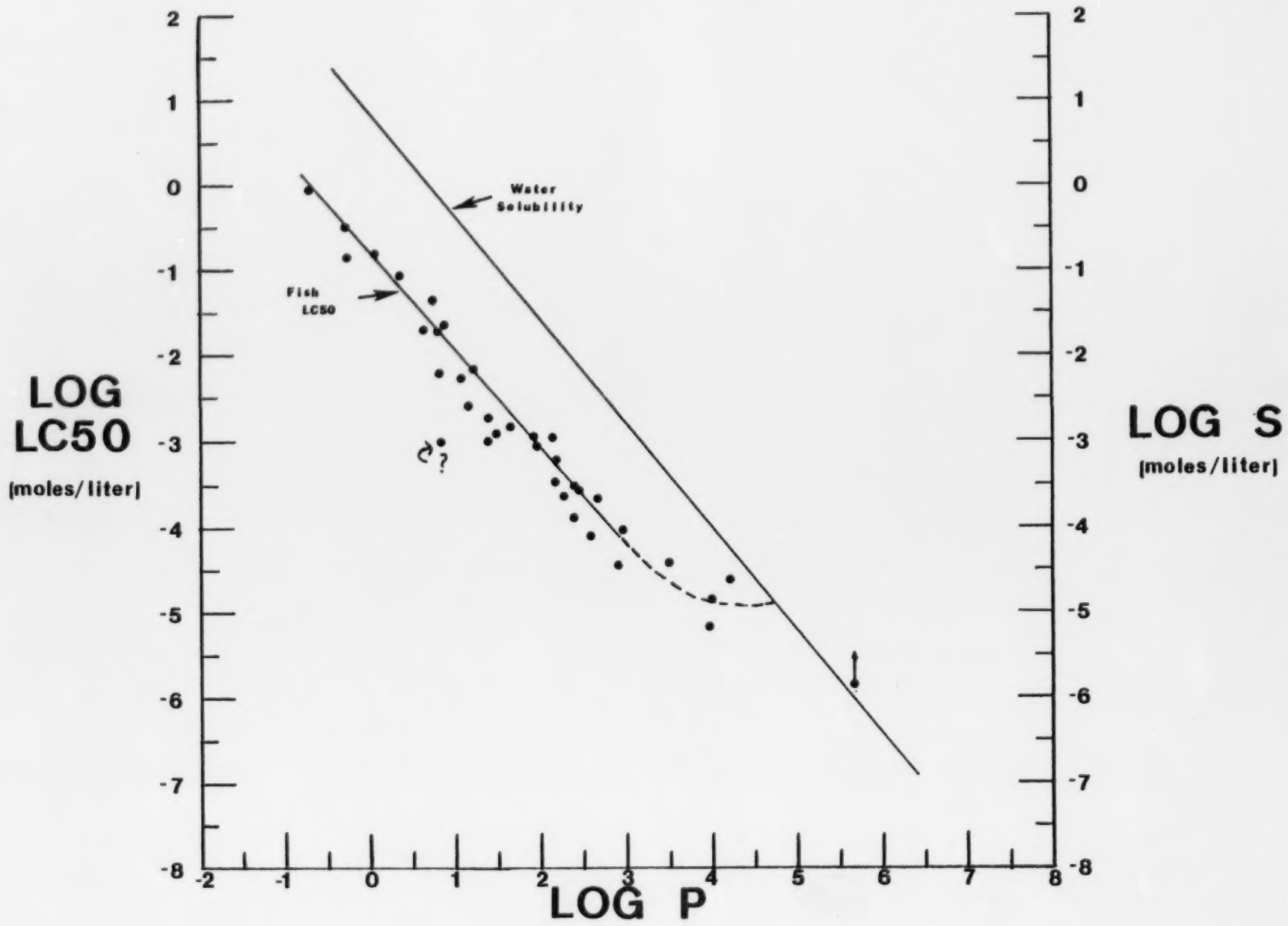


Figure 5. Relationship of Log LC50 and Log P for Narcotics

Two chemicals which were clearly below the "physical toxicity" line were 2-chloroethanol and 1-amino-2-propanol, both of which were much more toxic than would be predicted from this model. These chemicals (not plotted) both produced symptoms different from narcosis including bursting of the eyes from the fish. These chemicals are somewhat similar in that a single chemical substituent is located on the carbon adjacent to the carbon with the hydroxyl group. Tests are underway to evaluate the relationship between substitution patterns, metabolism, and lethal effects. This may be a good example of how physical toxicity is masked by an irreversible chemical toxicity which is observed in concentrations well below concentrations needed to produce narcosis.

V. MATHEMATICAL METHODS FOR QSAR

The development of Quantitative Structure Activity Relationships (QSAR) and pattern recognition techniques for drug design have grown rapidly since the Hansch and Free-Wilson methods and discriminant analysis (Jurs et al. 1969) were introduced. After computer time and statistical expertise became available, refinements and new methods were added to the organic chemist's analytical techniques. Free-Wilson, linear free-energy and bilinear models are separated from the other techniques only for convenience since all models and analyses share many common features. In discussion of each approach, the development of the mathematical model rather than chemical application should be stressed.

No matter how elegant the data treatment, the results will only be useful if the chemical and biological thinking is sound. Statistical significance will never replace intuition and logical thinking. Within a homologous series, regression analysis is the simplest method to show dependence of one variable on another. Usually the coefficients of regression analysis can be understood in terms of the independent parameters which allows the chemist to relate these parameters to biological mechanisms(s) and to make new predictions from the regression equations. If there is a choice of multiple regression methods, all possible subsets gives the best results. If nonlinear regression is available, the bilinear model best describes the apparent parabolic effect of biological activity on hydrophilic parameters.

When pattern recognition is used to describe large, diverse data sets, the first choice of methods should be factor and cluster analysis. Martin and Panas (1979) propose useful criteria and steps for using these methods in series design. Discriminant analysis can be useful when seeking broad

classification, but consideration should be given to the limitation and interpretation of results. Finally, all statistical analyses should include correlation tables of descriptor variables and development of equations (Ogina et al., 1980; Wu et al., 1980).

Free-Wilson Model. The Free-Wilson model assumes that each substituent makes an additive contribution to the biological response (BA). The model is expressed in equation 1,

$$BA = \sum_{ij} G_{ij} X_{ij} + \mu$$

where μ is an overall average and G_{ij} is the contribution of the substituent X_i in position j. With the additive assumption, each substituent effect at each binding site can be estimated by multiple regression analysis. The biological and mathematical assumptions and implications as well as modifications of the Free-Wilson model are summarized by Kubinyi and Kehrhahn (1976a,b). They have shown that the Fujita-Ban model has several advantages over the classical Free-Wilson and Cammarata models. In summary, the Fujita-Ban model requires fewer restrictions on the data matrix, gives estimates of the theoretical predictive activity value of the parent compound, shows lack of linear dependence and has a close mathematical relationship to the Hansch linear free energy model. Schaad and Hess (1977) reviewed the theory of linear equations and suggest ways to detect and circumvent linear dependence difficulties for all variations of the Free-Wilson model. Two general problems with the classical or modified Free-Wilson model is the inability to generalize biological mechanisms for a class of chemical compounds and the difficulty of preparing the data matrix before analysis.

Hansch Linear Free-Energy Model. The linear free-energy model has been used and modified numerous times. The model assumes a linear and additive

contribution from the hydrophobic, electronic, and steric properties of a drug or chemical compound on biological activity. Usually the log of the inverse of the biological response ($\log \frac{1}{C}$) is correlated to various physical-chemical parameters. The general form, equation 2,

$$\log \frac{1}{C} = a_i H_i + b_j E_j + c_k S_k + d$$

where the hydrophobic (H_i), electronic (E_j) and steric (S_k) effects are estimated by multiple linear regression coefficients a_i, b_j, and c_k respectively. The ease of calculation, availability of canned programs, and potential for describing biological mechanisms led to numerous publications of correlation studies.

Most problems that arise from using multiple linear regression are related to the choice and number of variables that are used to describe the biological response. The ratio of observations to variables should be greater than five. Topliss and Edwards (1979) have developed control charts for determining the probability of a chance correlation based on the number of observations and variables used in regression. These charts are based on the total number of variables that were tried as well as taking into account the collinearity of the descriptor variables. Unger and Hansch (1973) and Otto et al. (1979) have stressed the point that the investigator should not maximize the correlation coefficient at the cost of prediction by including more variables than can be supported by the data. Draper and Smith (1966) offer numerous statistical methods for evaluating the usefulness of regression coefficients and equations. The interdependence between physical parameters must also be kept to a minimum if correlations are going to describe different biological mechanisms of action. Craig (1971) and Hansch et al. (1973) have tested the interdependence of numerous substituent parameters by correlation analysis.

To further improve the interpretation of correlation studies, Unger and Hansch (1973) have presented criteria for selecting the "best equation". They have also listed the minimal information that needs to be included in all multiple linear regression analyses. Otto et al. (1979) have described a method that validates the final regression model. By applying the "leave-L-out" technique, they were able to show remarkable stability in their regression equations by keeping the number of variables to a minimum.

Recent papers by Iwamura (1980) and Wu et al. (1980) have included useful tables with a complete cross correlation of variables and the development of equations. Heymans et al. (1980) found a global steric effect in MR (molar refraction) but could have improved their interpretation by using an indicator variable for steric effects and an independent value for estimating the hydrophobic effect (Biagi et al., 1980; Wu et al., 1980). Both Biagi et al. and Wu et al. found that a steric indicator variable was superior than MR since MR was higly correlated with log P (Heymans et al., 1980 also reported this result). The indicator variable allowed these authors to test for specific steric effects that would otherwise be averaged out by MR. These analyses are simplified versions of the mixed model suggested by Kubinyi and Kehrhahn (1976) where a Fujita-Ban and Hansch model are combined. Biagi et al. (1980) also found that the regression technique of all possible subsets was superior to forward stepwise regression. One final problem with variable selection is the spread or standard deviation of the variable. Unless there is an adequate range for each variable used in the regression model then there is little chance of detecting the biological response to that variable (Martin and Panas, 1979).

Bilinear Model. The linear relationship between biological activity and the hydrophobic feature (log P) of a compound will not continue indefinitely.

Hansch and Fujita (1964) suggested a parabolic regression model to explain the nonlinear biological response to the increasing hydrophobic effect of a compound. To account for the strictly linear response for lower log P values, Kubinyi (1977) has proposed a generalized bilinear model, adapted from the McFarland (1970) probability model,

$$\log \frac{1}{C} = a \log P - b \log (\beta P + 1) + C$$

This generalized model generates curves with unsymmetrical, linear ascending and decending sides and a parabolic portion around the optimal value for log P. Through residual and statistical comparisons of 57 data sets, Kubinyi has shown the bilinear model superior to the parabolic model proposed by Hansch and Fujita. Kubinyi found situations when the parabolic model gives a better approximation of the biological response (few data points or small range for log P). This result can be attributed to the number of variables (4 versus 3) and calculation procedure (nonlinear regression).

Drug Design and Classification. The demonstrated relationship between biological activity and physical-chemical properties of chemical compounds has led to the search for methods that identify promising new leads. These methods can be separated into two categories, pattern recognition and non-mathematical searches.

(a) Pattern Recognition - Within the general field of pattern recognition, three categories can be found; linear discriminant analysis, cluster analysis and factor analysis. One of the first uses of discriminant analysis was to classify mass spectra (Jurs et al., 1969). Methods used for discriminant analysis include positive feedback functions (linear learning machines), linear discriminant functions, K-nearest neighbor, and threshold logical units. There are numerous examples that use discriminant analysis to

classify biological activity into two or three categories (Chu et al., 1975; Rasmussen et al., 1979; Henry and Block, 1979). All method of discriminant analysis, except K-nearest neighbor, attempt to divide a multidimensional data array by a hyperplane. K-nearest neighbor measures the distance between two or more positions in hyperspace (Albano et al., 1978). To improve the linear separation and reduce chance correlations, various data treatments were devised that verify the usefulness of the linear discriminant function (Kowalski and Bender, 1974; Weisel and Fasching, 1977). Before analysis data should be autoscaled, i.e. normalized to a zero mean and variance of one. After scaling, a subset can be used as a training set and another subset for verification of the discriminant function. Another common way to verify the discriminant analysis is by the Jackknife (leave-one-out) iterative procedure (Weisel and Fasching, 1977).

There are numerous factors that need to be considered before doing a discriminant analysis. Frank et al. (1965) and Morrison (1969) have summarized the difficulties in interpreting results from a "canned" statistical program. These considerations include equality of numbers and covariance matrices of observations between classes and the upward bias that results from verifying results with the same measurements used for classification. It is also true that the discriminant function may not be the "best" equation, since many solutions are possible for any data set. Gray (1976) found that the "noise" feature will influence the final discriminant function. Usually the ratio of observations to variables should be greater than three to avoid unstable (data dependent) discriminant functions.

A good example of how to properly use and discuss a discriminant analyses is given by Ogino et al. (1980). They have included tables of

variable correlations and the development of the linear discriminant function. These tables, which are similar to the tables used by Biagi et al. (1980), and Wu et al. (1980), are absolutely necessary for comparison and interpretation of the discriminant analysis.

Hansch and Unger (1973) were the first to use cluster analysis as a tool to find the most different analogue for a lead molecule. They used extrathermal substituent parameters as their factors and chose one compound from each cluster for synthesis. Cluster analysis uses the correlation matrix to find patterns of similar structure. There are numerous methods that group variables by the size and interdependence of their correlation coefficients. The outcome of any cluster analysis is dependent on the choice of information that goes into the analysis. Hopfinger et al. (1979) were not able to explain toxicity except in terms of molecular weight. This illustrates the need to choose factors that are orthogonal (not highly correlated) and descriptive of the chemical factors influencing toxicity. The indiscriminant inclusion of meaningless factors (unrelated to the biological property) will lower the predictive capability of the analysis and increase the possibility of getting a useless but successful classification.

The branch of statistical science called factor analysis has been used by psychologists since 1904. It wasn't until Cammarata and Menon (1976) suggested a classification scheme for handling large bodies of information that factor analysis was used to describe toxicity data. Their flowchart included factor analysis, discriminant anlaysis, cluster analysis, and finally correlation analysis to follow up a new lead compound. Several authors (Saxberg et al., 1978; Albano et al., 1978; Dunn and Wold, 1978) have compared factor analysis with discriminant analysis and have proposed a hierachy of classification ability with discriminant analysis on the lowest

level and factor analysis on the highest. Martin and Panas (1979) have proposed another strategy for exploring new "leads" in drug design. Based on a four part criteria, Martin and Panas suggest that factor analysis and cluster analysis will measure independence and uniqueness of each chemical compound.

Just as in any pattern recognition method, selection of variables is important. Martin and Panas (1979) state that the standard deviation of each variable be greater than one. They also suggest that the variables be tested for independence by factor analysis. Malinowski (1977a,b) has developed several methods for evaluating the usefulness of each factor by examining the error in the data matrix. Martin and Panas (1979) selected the number of factors used for classification by examining eigenvalues.

(b) Non-Mathematical Searches - Two non-mathematical approaches to optimize a lead chemical are the Topliss manual method (Topliss, 1977) and a Fibonacci search (Santara and Augary, 1975). Both methods select and rank a set of analogues by molecular properties usually found in a Hansch correlation study. The most active analogues are selected for synthesis based on the relationship between biological activity and physical parameters. The advantage of both methods is that neither requires a computer for analysis. Martin and Panas (1979) evaluated the Topliss manual method by factor and cluster analyses. They found that the strategy was useful but not ideal for finding the least number of compounds that covered the substituent space.

Perspective. The literature is replete with computer programs with intriguing names which are reported by their developers as QSAR methods. In general, the programs consist of one or more of the techniques found in systems such as BMDP which are available in many computing centers. In some

cases, such "canned" programs are tailored with respect to input/output routines to aid the analyst. The specific program used is dependent on the types of questions being asked.

This project will be evaluating many of the techniques discussed above using systematic data sets to demonstrate the relative contribution of each in addressing EPA needs.

VI. MOLECULAR DESCRIPTOR GENERATION

Attempts to predict biological or chemical activities from structural characteristics have involved well over 30 different characteristics which are loosely called molecular descriptors. Common descriptors, or predictors, of activity include log P, Hammet constants, Taft steric constants, molar volume, molar refractivity, CIDS keys, molecular orbital parameters, and connectivity indices. In many cases, each type of descriptor includes a family of parameters. Some descriptors can be selected to model certain activities because the descriptor is an important factor mechanistically in the process being modeled. However, many times the descriptors are selected on a statistical basis and their importance are rationalized as the model develops.

To keep this QSAR project directed toward the Agency's need for screening thousands of industrial chemicals, several constraints were placed on the selection of molecular descriptors. The first is to minimize the number of molecular descriptors which are used as independent variables in the QSAR models. The literature demonstrates clearly that there are few descriptors readily available for large numbers of industrial chemicals, and building models which require several descriptors further reduces the number of chemicals for which estimates of activity could be made from the models. The second is that molecular descriptors be selected where possible, which can be calculated directly from the structure of the chemicals. This would permit extensive computerized evaluations of the QSAR models on large numbers of chemicals.

A. Calculation of Log P

The discussion in previous sections of this report has demonstrated the usefulness of log P in estimating toxicity and BCF. Log P has been selected because it is a measure of the hydrophobic properties of the chemicals which is directly related to transport and binding in membranes. Moreover, the log P for thousands of chemicals can be calculated with adequate accuracy from structure. A project was developed with Dr. Al Leo and Peter Jurs to computerize the log P calculations. When a lack of funds prevented the timely initiation of the project, the Office of Toxic Substances funded the project. It is anticipated the first version of the program will be available to EPA in November, 1980. This project stands ready to evaluate the calculations on a subset of 15,000 industrial chemicals when the program is received.

B. Calculation of Molecular Connectivity Indices

Molecular connectivity indices are a family of constants which result from the application of graph theory to molecular topology. A review of the calculation and examples from use of these indices in physical and medicinal chemistry has been presented by Kier and Hall (1976) and is beyond the scope of this report. Briefly, the connectivity indices are calculated by subdividing each molecule into a series of subgraphs with the atoms as "vertices" and the bonds between them as "edges". The subgraphs are categorized as paths, clusters, path-clusters, and chains which are self-explanatory. The subgraphs are further divided into connectivity orders which are the number of edges. In general only the first four orders of indices have to be used in QSAR due, in part, to the difficulty in calculating the higher order indices for multicyclic molecules. For example,

endrin has in excess of 1,000 sixth order subgraphs, each of which must be analyzed individually and summed to give the individual sixth order, path, cluster, path-cluster, and chain indices.

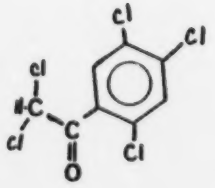
Kier and Hall further divided the indices into simple and valence-corrected indices. The simple indices are calculated as if all the atoms were carbon atoms but all bonding was the same as in the original molecule. The valence-corrected indices apply empirically derived correction factors for each type of heteroatom present in the molecule. The valence correction is intended to describe the variations in electron density around the heteroatom which may contribute to the reactivity of the molecule. The terminology employed is X to represent the simple indices and X^v to represent the valence-corrected indices. An additional superscript denotes the order such as ${}^4X^v$ being the fourth-order valence-corrected indices. If no subscripts are present, it is implied that the index is the sum of all the paths, clusters, etc. of that order. If individual indices are used, the subscripts "p", "c", "pc", and "Ch" denote the path, cluster, path-cluster, and chain indices, respectively. As an example, ${}^6X^v_{pc}$ represents a sixth-order valence-corrected path-cluster index.

A program has been developed by this project to calculate indices through the tenth-order indices for molecules including the multicyclic molecules such as mirex and endrin. The program calculates the valence-corrected indices, the simple indices which is the molecule reduced to a hydrocarbon of identical topology, and framework indices which is the molecule reduced to a saturated hydrocarbon of the same topology. An example of the data for molecular connectivity indices from this program is presented in Figure 6. The molecule considered is 2,2-dichloro-1-(2,4,5-trichlorophenyl) ethanone which has the structure indicated in the upper left. The

NAME: 2,2-DICHLORO-1-(2,4,5-TRICHLOROPHENYL)ETHANONE

C H CL O
8 3 5

INDEX VER.4 RUN ON 24-SEP-80



NUMBER OF SUBGRAPHS:

	ORDER										
TYPE	0	1	2	3	4	5	6	7	8	9	10
PATH	14	14	20	24	26	30	24	18	9	2	0
CL				6	0	4	0	2	0	1	0
CH				0	0	0	1	4	8	13	19
P/CL					20	39	87	133	184	208	179
TOTAL	14	14	20	30	46	73	112	157	201	224	198

VALENCE CORRECTED CONNECTIVITY INDICIES:

SUBGRAPH TYPE	ORDER										
	0	1	2	3	4	5	6	7	8	9	10
PATH	10.6596	5.5934	4.9769	3.0518	1.7933	1.3073	0.5904	0.2884	0.0889	0.0174	0.0000
CLUSTER				1.0708	0.0000	0.2733	0.0000	0.0365	0.0000	0.0171	0.0000
CHAIN				0.0000	0.0000	0.0000	0.0208	0.0857	0.1385	0.1357	0.1386
PATH/CLUSTER					1.8880	1.8790	2.8416	2.7798	2.5806	2.2048	1.6133
TOTALS	10.6596	5.5934	4.9769	4.1225	3.6813	3.4597	3.4528	3.1904	2.8080	2.3750	1.7519

SIMPLE CONNECTIVITY INDICIES: (REDUCED TO HYDROCARBON)

SUBGRAPH TYPE	ORDER										
	0	1	2	3	4	5	6	7	8	9	10
PATH	9.9392	5.2016	4.4313	2.8836	1.6175	1.1100	0.5093	0.2335	0.0703	0.0120	0.0000
CLUSTER				0.8959	0.0000	0.2565	0.0000	0.0465	0.0000	0.0170	0.0000
CHAIN				0.0000	0.0000	0.0000	0.0208	0.0729	0.1071	0.1065	0.1139
PATH/CLUSTER					1.7325	1.7049	2.4253	2.3179	2.1009	1.6976	1.1542
TOTALS	9.9392	5.2016	4.4313	3.7795	3.3500	3.0715	2.9554	2.6707	2.2784	1.8351	1.2681

FRAMEWORK CONNECTIVITY INDICES: (REDUCED TO SATURATED HYDROCARBON)

SUBGRAPH TYPE	ORDER										
	0	1	2	3	4	5	6	7	8	9	10
PATH	10.8783	6.4304	6.1702	4.8167	3.1147	2.3869	1.2670	0.6225	0.2091	0.0370	0.0000
CLUSTER				1.3696	0.0000	0.5339	0.0000	0.1341	0.0000	0.0556	0.0000
CHAIN				0.0000	0.0000	0.0000	0.0556	0.1987	0.3135	0.3591	0.4061
PATH/CLUSTER					3.1991	3.7001	5.8225	6.2653	6.2663	5.4691	3.9779
TOTALS	10.8783	6.4304	6.1702	6.1857	6.3138	6.6209	7.1450	7.2207	6.7889	5.9208	4.3840

Figure 6. Connectivity Indices for 2,2-dichloro-1-(2,4,5-trichlorophenyl) ethanone

Reproduced from best available copy.

table in the upper right summarizes the number of subgraphs in the molecule. The molecule has 14 atoms as evidenced by the 14 zero-order subgraphs. The 6 third-order clusters are located at the six carbon atoms which are substituted. The phenyl ring obviously accounts for the sixth-order chain. The lower half of Figure 7 presents the individual and summed indices for this molecule.

Dr. David Weininger and Brad Greenwood at ERL-D have a complete package of programs called INDEX which allow structures of chemicals to be entered either manually on a remote TTY or with a graphics terminal, to check the molecule for proper valencies to calculate a complete set of up to tenth order indices, file the files along with other data on chemicals for input to programs such as BMDP or other statistical packages. Approximately 800 chemicals which are part of one of the QSAR data bases have been entered into the system.

The INDEX system is a group of programs, command files, and structured data files which currently allow:

-interactive graphic molecular structure specification,

-verification of "foreign" files of molecular structures,

-hard- or CRT-graphic display of chemical structures on files,

-computation of connectivity indices of the 0^{th} to 10^{th} orders,

-storage of connectivity indices for up to 20,000 chemicals in compressed form (the master file),

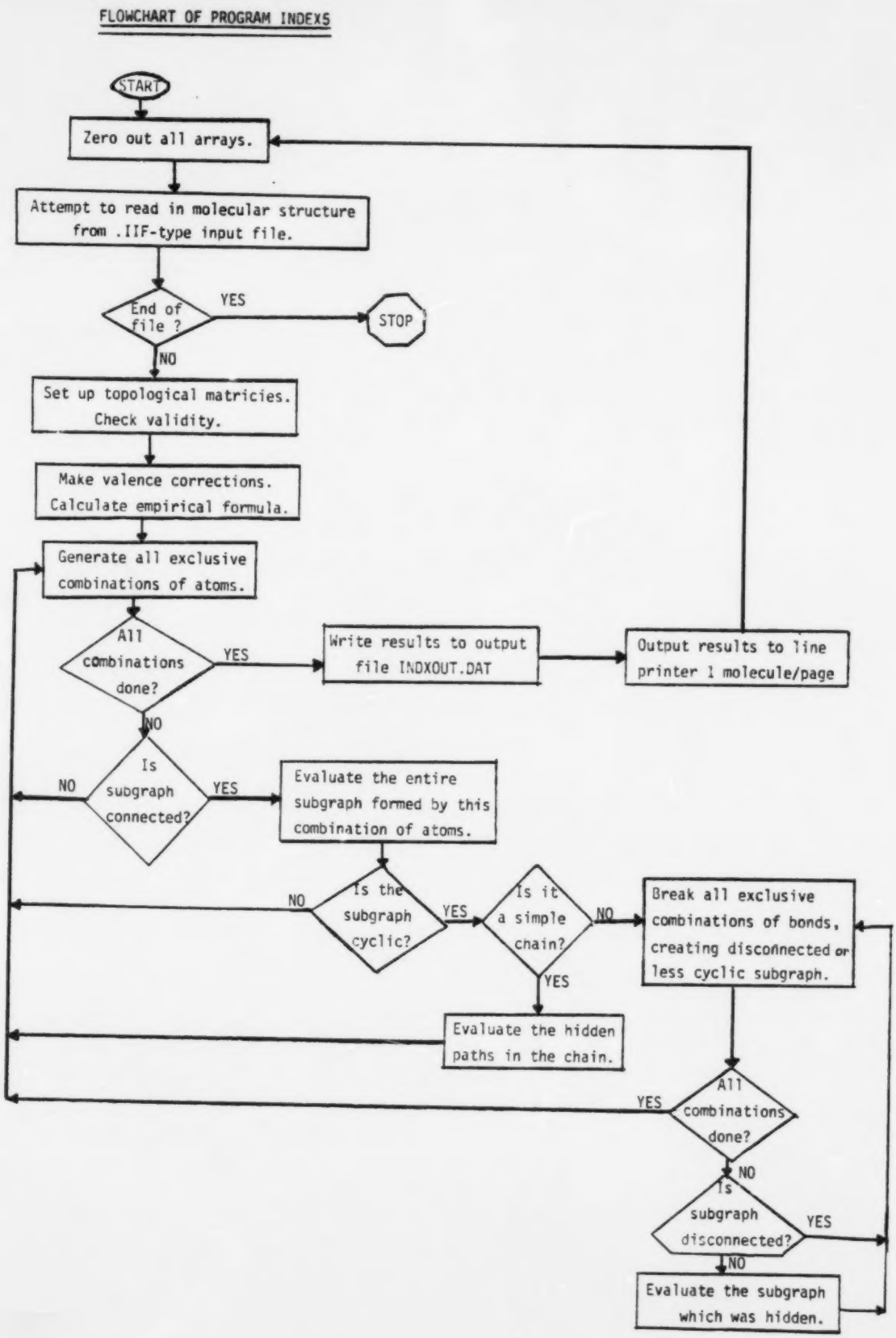
-a variety of options for retrieval from the master file,

-generation of master file table of contents ordered by location or by empirical formula,

-automatic detection of duplicate entries on master file,

-encoding an additional parameter along with the connectivity indices of selected compounds for subsequent BMDP analysis,

Figure 7. Flow-Chart of Program INDEX 5



-locating the nearest neighbor (and most distant) of the master file compounds in connectivity factor space,

-and several other utility operations.

Additional capabilities expected to be implemented by 11/80 include automatic master file backup and a routine for determining if a given molecular structure falls inside or outside the hypersolid defined by a group of molecules in factor space. A general description of the INDEX programs and files follows.

Creating and Handling INDEX Input Files. The first step in the process of molecular structure description is unambiguously specifying a molecular structure. The format used in INDEX is an extended, linearized connectivity matrix; files containing lists of molecular structures in this format can be identified by the file extension ".IIF" (Index Input Format). Although not yet tested, the translation of "foreign" topological matrixes to IIF format should be straightforward.

Program MOSS (MOlecular Structure Specification) is intended to be the primary means of initially specifying molecular structures. MOSS is an interactive graphics program designed to be run on a Tektronics 4025 Graphics Terminal. The program outputs the structures of graphically displayed molecules in a verified IIF format file. Verification includes checking atom types and valence validity and extended topological matrix symmetry. The user may specify the entire structure or modify any structure stored on an IIF file. It is up to the MOSS user to verify that the correct name is associated with a molecular structure.

Creating and Filing Connectivity Indices. The program INDEX5 reads molecular structures from an IIF file, computes molecular connectivity indices, prints tabulated results, and writes the results in the file

INDXOUT.DAT. The program uses a general algorithm for connectivity index computation; the following limitations are imposed by available valence corrections, computation time, and storage effectiveness considerations:

-allowable non-H atoms: C, N, O, S, Br, Cl, I,

-maximum number of non-H atoms per molecule = 50,

-highest order connectivity index computed = 10, and

-highest allowable cyclic order of 11-atom-subgraph = 5.

The INDXOUT.DAT file contains the 100 character name, the coded empirical formula, and the following information for each of the 34 possible subgraph types of orders 0-10: the number of unique subgraphs, the valence-corrected CI, the simple CI, and the framework CI.

The formatted sequential nature of the INDXOUT.DAT file makes it inefficient with respect to disk storage space. Therefore, the information contained in INDXOUT.DAT is transferred to a compressed master file (MINDEX.DAT) by the program LOADEM. A successful INDEX5 run is normally followed by a LOADEM run. This two step process is used to ensure the integrity of MINDEX.DAT in the event of a fatal error during INDEX5 execution.

Accessing Connectivity Indices on the Master File. All user access of the master file must be done via program control because the file contents are written in binary. The master file, MINDEX.DAT, is set up for "direct access" which allows any record to be directly retrieved by specifying a location.

Three programs are specifically designed to retrieve master file contents for specified chemicals. Programs GETEM and FILEM are essentially identical programs except for the output file: GETEM prints the retrieved entries in their original tabular form, FILEM writes entries into the file,

SHMIND.DAT. Both programs require that the master file locations be specified and provide a means of locating an entry by empirical formula. Any number of entries can be retrieved during a single GETEM or FILEM run.

A third retrieval program, PREPEM, is similar to FILEM but also has these additional features:

-The input file can be the master file, the file INDXOUT.DAT, and/or any other file in the INDXOUT.DAT format.

-An additional molecular descriptor can be specified (eg., "LOG P") for each compound included in the output file.

-The molecular descriptor name and values are stored in the last 20 characters of the 100 character name.

-Any output file name may be specified. This file can be directly used by any BMDP program as a data file.

The program INTDIST is used to find which compounds in the master file are most and least distant from a given chemical in the 34-dimensional valence corrected connectivity space. The user can select the "given" compound from the master file on any INDXOUT.DAT-type file. INTDIST will list the nearest and farthest compounds in the master file, automatically discounting duplicates.

Operation of the INDEX5 Program. INDEX5 is the fifth version of a series of programs which compute molecular connectivity indices for organic molecules. As discussed above, INDEX5 uses a general algorithm for connectivity index computation. A flowchart of INDEX5 is presented in Figure 8. Each molecular structure in an INDEX Input Format (IIF) file is transformed initially into two topological matrices: the simple connection matrix which contains only 1's and 0's (T), and an extended matrix which

contains the topological valence for each connection (PT). Valences are then calculated based on both topological valence (number of connections to a vertex) and elemental valence corrections.

The INDEX5 algorithm is vertex oriented. All exclusive combinations of up to 11 atoms are generated and the connectivity index contributions are evaluated for the subgraphs which include all atoms in the combination only. The decision to evaluate up to 10th order connectivity indices requires that up to 11 atoms must be considered at a time. An efficient pseudo-recursive (stack based) algorithm is used to generate the combinations. CONECT's high efficiency is obtained by automatically suppressing generation of disconnected combinations.

Each connected combination generated by CONECT is evaluated by the first level routine EVAL. The entire subgraph (represented by all atoms and bonds in the combination) is evaluated. The term "evaluate" is used here to mean "determining subgraph type and order, incrementing the ennumerative counters, computing the subgraph index contributions, and adding these contributions to the appropriate index accumulators". If the subgraph is cyclic, it will contain a series of hidden subgraphs which must be evaluated. In the case of a simple chain of N atoms (all topological valences = 2), EVAL evaluates the N hidden paths of N-1 order. In more complex cases, EVAL determines the number of cycles in the subgraph and passes the hidden subgraph evaluation to the routine CYCLIC.

The subroutine CYCLIC evaluates hidden subgraphs by using the following theorem: "If an edge is removed from a graph which is K cyclic, the resultant graph is either (K-1) cyclic or disconnected". Combinations of bonds (edges) are broken such that not only are the broken bond combinations

exclusive, but so are the combinations of remaining bonds. Subroutine CYCLIC is capable of breaking up to five edges of a subgraph simultaneously enabling analysis of 5-cycle, 11-atom subgraphs. The above algorithm circumvents the difficulties associated with multiple cyclic subgraphs; by breaking up to K bonds of a K cyclic subgraph one is assured that all possible hidden subgraphs will be uncovered (whether or not they are in themselves cyclic). Responsibility for evaluation of the generated subgraphs is passed to EVAL2.

Cyclic subgraphs generated by edge removal present a further difficulty. It is not possible to directly determine from the topological valences whether the subgraph is connected or whether it represents two connected subgraphs which are disconnected from each other. A special routine, ICHECK, is designed to cope with this problem. ICHECK uses a "route-finding" algorithm to ensure that there is a route from every vertex to every other.

Subroutine EVAL2 is similar to EVAL except that EVAL2 does not need to determine cyclic order and does need to use ICHECK to determine connectivity.

After all evaluations have been made for the combination of atoms generated by CONECT, control is passed back to CONECT and the next combination is generated. After all exclusive connected combinations of atoms have been evaluated, CONECT passes control back to the main program for output of results.

An output routine, OUTPUT, writes the results to the file INDXOUT.DAT. This file is a formatted sequential file which contains: the molecule's name and empirical formula, the data run, enumeration of each subgraph type and order, and valence corrected, simple, and framework connectivity indices for orders 0 to 10. OUTPUT also sends this information to the line printer in a tabular form which requires one page per molecule.

Glossary of Programs and Files. All important programs, command files, and data files used in the INDEX system are listed below with brief explanations. Alphabetical order is used with the file extensions having priority. In general, file types can be identified by their file extensions:

.BIG - an IIF file contining "big" molecules only

.BKP - master file backup

.CMD - command file

.DAT - general data file

.FTN - Fortran code (used here for subroutines)

.IIF - INDEX input file

.MED - an IIF file containing "medium" molecules only

.SML - an IIF file containing "small" molecules only

.TOC - at table of contents file

.TSK - an executable task image of a program

Note that many programs are listed which will never be run directly and some files are listed which are transparent to the user. Almost all user executions will use only .CMD and .TSK listings below. The program INDEX5 and its subroutines are not described in the glossary (see "Operation of the INDEX5 Program", below).

xxxxxx.BIG - An IIF-type file written by the program SPLITTER which contains only molecules with 23 atoms or more.

xxxxx.BKP - A backup file for the master file which was created on the date "xxxxx" (e.g. OCT17.BKP). This file is in the INDXOUT.DAT format.

DFILE.CMD - Command for duplicate search of the master file.

DSORT.CMD - Command for a PDS sort used in DFILE.CMD.

PSTOC.CMD - Command to print table of contents of the master file ordered by location or by molecular formula. The master file can also be output to the line printer in tabular form with this command.

TOC.CMD - Command for a PDS sort used in PSTOC.CMD.

SHMIND.DAT - An INDXOUT-type file containing selected entries from the master file written by the program FILEM.

DISCOM.FTN - This subroutine takes a given compound with its connectivity indices from the program INDIST, then searches the master file computing the Euclidean distance for each compound in a connectivity 34-space. The most and least distant compounds are then output to the user's terminal.

MFORM.FTN - This routine outputs molecular formulas to the line printer.

PUTOUT.FTN - This routine takes an entry from the INDEX5 master file and outputs it to the line printer in table form.

RITE47.FTN - This routine writes a sequential file with parameters in the title field of each record retrieved from a user selected sequential file in INDXOUT.DAT format.

RITEM.FTN - This routine retrieves an entry from the master file and writes it to a sequential file on logical unit 3.

RRITEM.FTN - This routine writes a sequential file with parameters in the title field of each record retrieved from the INDEX5 master file.

SEQDIST.FTN - See write up for DISCOM.FTN. The difference is that here the initial compound is retrieved from a sequential file in INDXOUT.DAT format rather than from the INDEX5 master file.

xxxxxx.IIF - A file in the INDEX input format. Contains coded molecular structures. (e.g. PHENOLS.DAT)

MASTERxxx.IIF - An IIF file which is saved. Normally a number of .IIF files are appended together to form the type of file (eg. MASTER6.IIF).

xxxxxx.MED - See xxxxxx.BIG; .MED contains only molecules with 12-22 atoms per molecule.

xxxxxx.SML - See xxxxxx.BIG: .SML contains only molecules with 11 or fewer atoms per molecule.

TOC.TOC - A transparent temporary file for input to a PDS sort (TOC.CMD) used in the table of contents programs (PTOC.CMD).

STOC.TOC - A sorted transparent temporary file output from TOC.CMD in the table of contents programs.

BURP.TSK - This program initializes an available space list for the master filing system of INDEX5.

CHECKERS.TSK - This program checks the validity of molecular structure entries in an IIF file. The program is designed to be run on a Tektronics 4010 terminal. Each molecule is drawn; hard copies are optionally available. "Wheat" and "chaff" filing optionally available.

CINAMES.TSK - This program prints the names of molecules on an INDXOUT.DAT-type file.

DFILEM.FTN - This program is part of the duplicate entry search of the INDEX5 master file. It writes the entire master into a temporary sequential file to be used in a PDS SORT command (DSORT.CMD).

DSFILE.FTN - This program uses the files created by DFILEM and DSORT to locate duplicate entries in the INDEX5 master file.

FILEM.TSK - This program writes a sequential file in INDXOUT.DAT format from sequentially or randomly retrieved compounds from the master file.

GETEM.TSK - This program retrieves entries in the INDEX5 master file by location and outputs them to the line printer in table form.

HIBURP.TSK - This program initializes the hash table for the INDEX5 master file.

IIFCOUNT.TSK - This program counts the number of molecules in an IIF file and prints a frequency histogram of the number of atoms per molecule.

IIFNAMES.TSK - This program will print a list of the names of molecular structures in an IIF file.

INDIST.TSK - This program takes a user selected compound from the INDEX5 master file or a sequentia' file in INDXOUT.DAT format and compares it to every non-garbage, non-duplicate entry in the master file and computes the most and least distant compounds with respect to the connectivity indices in 34-space.

INTMIND.TSK - This program initializes the master file MINDEX.DAT for the INDEX5 filing system.

LOADEM.TSK - This program uses the output file (INDXOUT.DAT) created by INDEX5. After the data in the file has been verified, this program is run to insert the entries into the master file.

MOSS.TSK - This is an interactive Molecular Structure Specification program which outputs a verified IIF-type file. The user can specify entire structures or modify any existing structure already on an IIF file. Designed to be run on a Teletronics 4025 terminal.

PREPEM.TSK - This program creates a user named output file with a parameter name and value for each logical record in the title field for the compound. Input can be taken from the master file or any file in INDXOUT.DAT format.

PSTOC.TSK - This program prints out a table of contents for the INDEX5 master file sorted by location or molecular formula. It can also output the entire master file to the line printer in tabular form.

PTOC.TSK - This program creates a temporary sequential file to be used for input for TOC.CMD and PSTOC.TSK.

SPLITTER.TSK - This program splits an IIF file into three files (.SML, .MED, and .BIG) based on the number of atoms per molecule.

WIPEM.TSK - This program deletes entries from the master file given the entry location.

Despite the immediate usefulness of some connectivity indices and the insignificant cost of calculating them for large numbers of chemicals, there are two problems which must be solved before a complete set can be prepared. The first is that the computer program operates from a connection table which must be generated or purchased. Even if 10 structures/hour could be encoded into the system, over 500 working days would be needed to input the TSCA inventory in the connectivity program. An alternative would be to purchase the

connection tables from existing files; however funds to accomplish this have not been available to this project. The second limitation is that adequate valence correction factors have not been determined for all heteroatoms. Consequently, valence-corrected indices cannot yet be calculated for organophosphorus chemicals. This work can proceed if funding is available.

One of the potential uses of the connectivity indices in addition to single preditions is the development of a "similarity index" for chemicals in the n dimensional structure space. The purpose of this project is to provide the Agency with the capability to "indict" a chemical because of its similarity to hazardous chemicals for which data are available. The "similarity index" may evolve into a ranking system for testing priorities. For example, each chemical structure can be represented by a single point in n dimensional (often more than 30 dimensions) structure space. Chemicals with similar properties form clusters in structure space. The distance of any new chemical structure from a given cluster can be calculated to determine statistically whether the new structure is "inside" or "outside" the cluster and, if outside, how it is distributed between clusters representing contrasting environmental behaviors. This part of the QSAR project is evolving as a "back-burner" project until funds can be redirected from testing to develop it more rapidly.

For a more detailed account on possible uses of the indices, the reader is referred to Kier and Hall (1976). As a summary, it has been shown that the boiling point of classes of chemicals can be estimated from the following equations:

alkanes $\quad b.p. = 55.69{}^1X + 4.708{}^4X_{pc} - 96.13 \qquad r = 0.9969$

alcohols $\quad b.p. = 191.83{}^1X - 155.23{}^1X - 17.99{}^3X_p + 21.74{}^3X^v_p - 2.74{}^3X_c - 29.05 \qquad r = 0.99$

alkyl halides $\quad b.p. = 36.36{}^1X^v + 16.36{}^3X_p + 2.869{}^3X^v_c - 31.037 \qquad R = 0.9982$

The water solubility of alcohols can be estimated from

$$\ln S = 9.417 - 11.266{}^1X + 8.643{}^1X^v \qquad r = 0.9945$$

Log P can be estimated from a similar series of equations such as for aliphatic carboxylic acids

$$\text{Log } P = -0.859 + 1.615{}^1X^v - 0.550{}^1X \qquad r = 0.9979$$

Similar equations have been developed for many biological activity endpoints such as narcosis, conversion of cytochrome P-450 to P-420 in rat livers, thymidine phosphorylase inhibition, microbial inhibition, vapor toxicities, and potency of mutagens in the Ames test.

VII. QSAR INFORMATION SYSTEM

An information storage and retrieval system was first begun as part of the structure-activity studies. It was clear that any computer modeling study would require the storage, retrieval and manipulation of large amounts of physical, chemical, and biological data. Efforts to determine the availability of a commercial or public information system capable of meeting the needs of the project met with little success and work was begun on the design of a local interactive information storage and retrieval system and a corresponding data base.

The first data base contained only 248 compounds with the following data elements: a local identification number, Chemical Abstracts registry number, molecular weight, molecular formula, freezing point, boiling point, vapor pressure, solubility in water, Wiswesser line notation (WLN), logarithm of the partition coefficient (log P), gas chromatographic retention data, bioconcentration factors, acute aquatic toxicity data (LC50), and acid dissocation constant data.

Initially the information storage and retrieval system was quite primitive as searches could only be carried out for the following: one local identification number, one registry number, a specific or ranged molecular weight, a ranged, partial or complete molecular formula, or a specific or ranged log P value. The commands available at that time were:

SEARCH	COMMAND
DISPLAY	END
CLEAR	

In the next phase of system development an additional 2,000 compounds were added to the data base along with new data elements such as maximum allowable toxic concentration and 8th and 9th Chemical Abstracts Collective

Index, systematic, common, and trade names. Many improvements were made in searching capabilities, for example, an exact name search was added and all properties could be "inventoried". Numerous new commands were added:

INTERSECT	SHOW
MERGE	HELP
LIST	STATUS
OUTPUT	COMMENT
INFORM	STOP

During this development the acronym ISHOW (Information System for Hazardous Organics in Water) was coined, a User's Manual prepared, and the system was made available to the public.

Additional data were added to the data base and considerable effort was expended in continued software modification to provide for rapid searching compatible with efficient storage. A major effort was the development of a software manual along with complete documentation of the computer software.

ISHOW is currently undergoing another major change to test methods of evaluating QSAR on large numbers of industrial chemicals. Using information obtained from the Office of Toxic Substances, the ISHOW data base is being expanded to contain an additional 12,000-13,000 compounds as a training set of data. Through the support of EPA Region V an expanded ISHOW will contain compounds manufactured or used in the Great Lakes watershed or EPA Region V. Corresponding information on manufacturers or users of these compounds in the Great Lakes is being added, along with new data elements which include biochemical oxygen demand, chronic toxicity (aquatic), and the Kovats index for gas chromatography data. Software to store and search these parameters is also being developed.

The purposes of this part of the project are twofold. First, this project is attempting to make data concerning the biological and chemical activities as well as key molecular descriptors available to many researchers with the assumption that this field of study will progress more rapidly. Second, it is attempting to develop and test predictive techniques as they are developed on large numbers of chemicals such as those found in the TSCA inventory. Such evaluations are judged to be essential if QSAR is to become an acceptable screening technique in evaluating the environmental efficacy of organic chemicals on a major scale.

VIII. ACTIVITY REPORTS

A. Compilation of the Aquatic Toxicity Data Base

A review of the aquatic toxicity data in the literature indicated early in the study that the amount of data as well as the quality of comparable data precluded serous QSAR modeling attempts. This conclusion was a major factor in the decision to generate a systematic set of toxicity data with aquatic organisms. In the event that data in the literature can be used in the QSAR models after the patterns begin to emerge from the testing program, the QSAR project is compiling an aquatic toxicity data base.

The data are being compiled from manual and computerized literature searches using the compiling form listed in Figure 8. The form was designed with the cooperation of ASTM and OTS and is intended to meet the needs of several computerized data bases. The data will be loaded into a computer file at ERL-Duluth from which subsets of the data will be sent on magnetic tapes to the ISHOW and HEEDA developers according to their respective needs.

To date, information on the toxicity of approxiately 700 chemicals has been compiled. The data include tests on approximately 75 freshwater and marine organisms. Acute tests make up approximately 675 of the tests and chronic tests constitute about 75 of the tests. Fathead minnows, rainbow trout, guppies, bluegills, daphnids, scuds, and stoneflies were most frequently tested. In addition to the toxicity data, approximately 50 tests have been reported for bioconcentration. At present ISHOW contains test information on 350 chemicals and information on approximately 150 chemicals is ready to be loaded into ISHOW. Tests have been performed on about 200 chemicals and/or mixtures for which Chemical Abstract Service registry numbers have not yet been located.

Figure 8. Aquatic Toxicity Data Compiling Form

CHEMICAL NAME _____

CHEMICAL ABSTRACT REGISTRY NUMBER (CAS #) _____

TEST CHEMICAL		END EFFECT		TEST CONDITIONS									
Chemical Grade	Source	Effect Endpoint	Conc. (Mg/l)	Species	Exposure Time	Lifestage	Exposure Type *	Conc.**	Temp	Hardness	Alkalinity	Salinity	pH
REFERENCE													
REFERENCE													
REFERENCE													
REFERENCE													

* S-Static, FT - Flowthrough

** M - Measured, N - Nominal

B. Estimation of the Bioconcentration Factor

The measurement and estimation of the bioconcentration factor (BCF) in fish have been reviewed by Veith et al. (1979, 1980). It has been shown that the BCF can be estimated using the equations:

$$\log BCF = 0.85 \log P - 0.70$$

or

$$\log BCF = 0.76 \log P - 0.23$$

depending on the data base selected. The accuracy of predictions made with these QSAR models is limited by the inherent variances in the BCF test and the variation of species tested in the data base. The equations were intended for an estimation of the bioconcentration potential within an order of magnitude for screening purposes. Chemicals for which metabolic routes not reflected in the log P are possible may give BCF values lower than the bioconcentration potential predicted by log P. For the majority of chemicals tested the present predictive capability can estimate the BCF within a factor of three.

Improvements in these predictive capabilities are being sought along several fronts. The first is the continuing effort to compile BCF data into a QSAR computer file which is interfaced with automatic plotting and regression programs. This will permit frequent updating of the QSAR for BCF at the request of the EPA program offices.

The second project is a re-analysis of BCF test data to improve the calculated BCF values reported for individual chemicals using methods recommended by OTS or ASTM. This effort began with a study of the BCF data for trichlorobenzene, DDE, and hexachlorobenzene as part of an EPA round-robin test validation program for OTS. A complete analysis of this data will be reported elsewhere and summarized in subsequent progress reports of this project.

The third project is aimed at uncovering the variation of the BCF among several species tested. In an attempt to relate the BCF of specific chemicals between fathead minnows, bluegills, rainbow trout, etc., a simplified bioenergetics model is being tested to determine the extent to which the differences in observed BCF between species is dependent on different metabolic activities of the species. Exposure equipment has been constructed to measure the BCF and the oxygen consumption of fish simultaneously. Since oxygen consumption may reflect the amount of water pumped across the gills, it is hypothesized that this value may "normalize" the BCF value for different species. This study will be discussed in more detail when appropriate.

C. Estimation of Water Solubility

The solubility of many organic chemicals has been estimted from the log P by numerous investigators. Hansch et al. (1968) presented relationships for 9 classes of organic chemicals showing that linear inverse relationships exist over approximately six orders of magnitude of water solubility as shown in Figure 9. If the data are separated into chemical classes, a series of equations estimating solubility from log P are available with a correlation coefficient greater than 0.96 for most classes and greater than 0.93 for all classes.

Because of the importance of water solubility in defining the aquatic toxicity testing regime as discussed previously and in the distribution of chemicals in the environment, a computer data base of solubility data has been established at ERL-D. Although there are insufficient funds at this time to generate water solubility data in this project, an effort is being made to update the data base from literature reviews. Subsequent reports

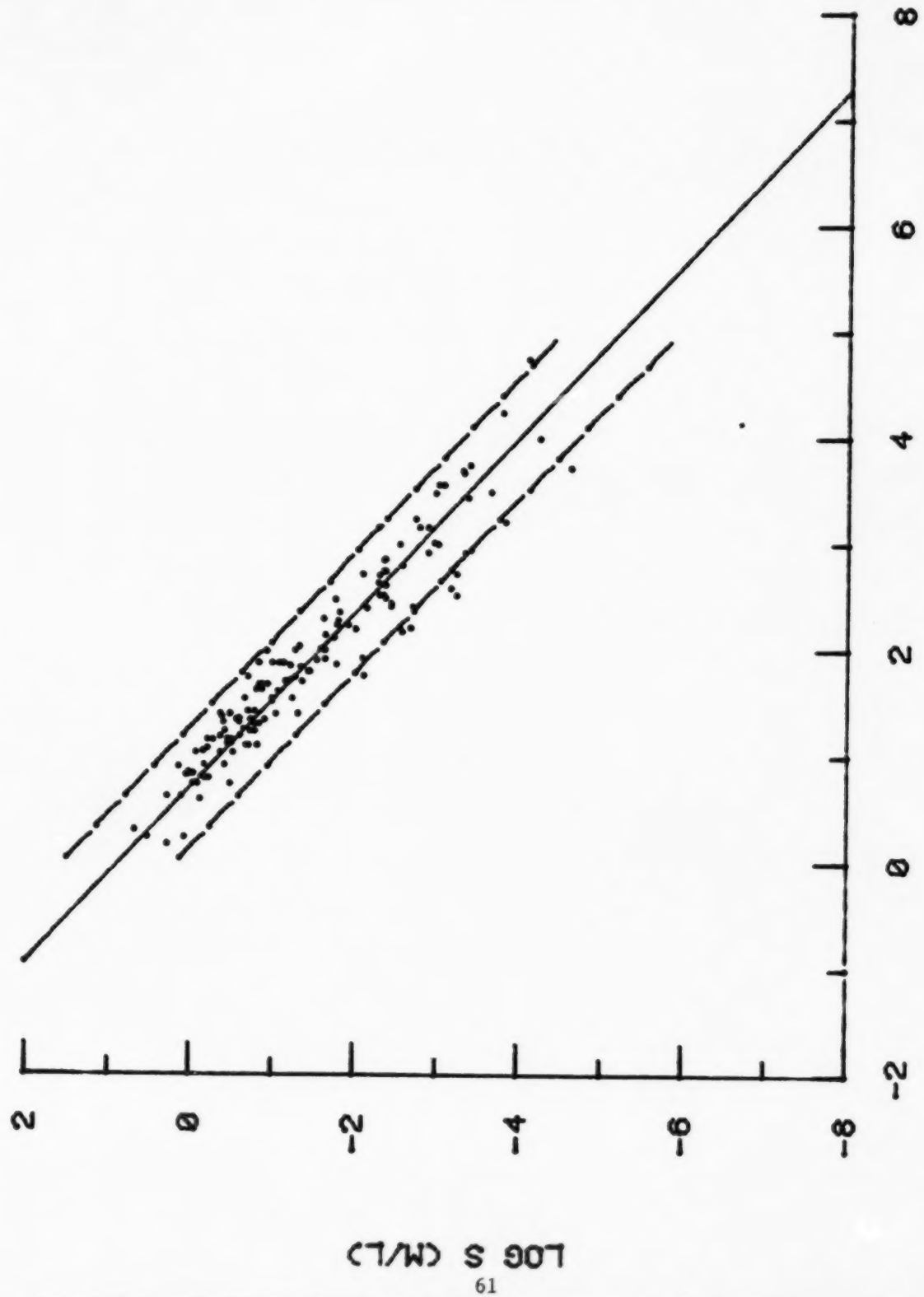


Figure 9. Relationship between Log P and Water Solubility

from this project will provide evaluations of the state-of-the-art in estimating water solubility by QSAR.

IX. EVALUATION OF RAPID SCREENING TESTS FOR QSAR

In the event that the QSAR analysis of aquatic toxicity data leads to the need to test more chemical classes than is feasible using the four day acute test, this project began evaluating several more rapid tests for the reliability in estimating relative toxicity of chemicals. The tests initially selected were the residual oxygen bioassay, the Microtox® bacterial bioluminescence bioassay, and the fluorescence algal stress bioassay.

A. Residual Oxygen Bioassay

The residual oxygen bioassay is a 6-8 hour test performed in sealed bottles. Briefly, about 3 gm of fish are placed into each of a series of BOD bottles containing water saturated with oxygen and varying concentrations of the test chemical. The fish die quickly at lethal concentrations of the chemical or after longer periods at sub-lethal concentrations due to oxygen depletion. The oxygen is measured in each bottle as the last fish dies and the residual oxygen concentration is plotted against the concentration of the chemical.

At high concentrations, the fish are killed quickly by the chemical before the oxygen can be reduced. In control exposures, the residual oxygen remaining for fathead minnows is approximately 0.5 mg/L. The concentration of chemical at which the residual oxygen rises above that of the control residual is the lethal threshold for the chemical. This study is an evaluation of the relationship between the lethal threshold and the 96 hour LC50 for fathead minnows.

Tests were conducted in 300-ml BOD bottles containing the desired concentrations of test chemicals. Stock solutions were prepared by either of two ways. For some solutions the calculated amount of chemical (solid or

liquid) was weighed out, transferred to a 2-liter volumetric flask, and diluted to volume; aliquots of this stock solution were then transferred to the BOD test bottles and dilution water was added to obtain the desired test concentration. For other solutions the stock solution was prepared by taking 2 liters of solution from a saturator, analyzing this solution, and then preparing the desired concentrations from aliquots of the stock solution as before. The dilution water used in all tests was Lake Superior water warmed to room temperature (20-22°C).

Fathead minnows (Pimephales promelas) from the Environmental Research Laboratory-Duluth stock were used as the test fish; these fish ranged from 30-day-old fry to adults. Fish were held in the same water at the same temperature as that used for test dilution water. During holding, the fish were fed either frozen brine shrimp (San Francisco Bay Brand, Inc., Newwark, CA) or a commercial salmon ration (Glencoe Mills, Inc., Glencoe, MN) twice daily. Fish were not fed for at least 16 hr prior to testing nor during testing.

Tests were conducted by preparing the test solutions as described above, usually using 9-11 test concentrations plus a control (dilution water only) for each test. Fish were weighed to obtain a 3 to 3.5-g loading for each BOD bottle, were rapidly placed in the test solution, and the bottles were immediately capped. When all the fish at a given concentration died, the dissolved oxygen concentration and pH of the test solution were measured, and the number of test fish was counted.

Dissolved oxygen concentration was determined using a Beckman 0260 oxygen analyzer calibrated by a Winkler titration (APHA et al., 1976) on water from a holding tank; pH was determined with a Corning Model 12 meter.

Averaging of pH was accomplished using a computer program which converts pH values to their corresponding hydrogen ion concentrations, averages these, then reconverts the average to pH units.

Data Treatment. The dissolved oxygen concentration and toxicant concentration test data for each test were analyzed by computer to determine the lethal threshold concentration. The lethal threshold endpoint was determined by means of the computer program HOCK(ey-stick) which fits the data to a model for the bioassay and used "hockey stick" statistics to estimate the endpoint confidence intervals. This program was developed to offer maximum input/output flexibility.

Version I enables the user to input data directly into the HOCK program and receive output from the line printer. This is for single time usage. Version II uses a data file to store information for multiple use. Version III gives examples of data modification and terminal output and can be used with either a data file or terminal data input.

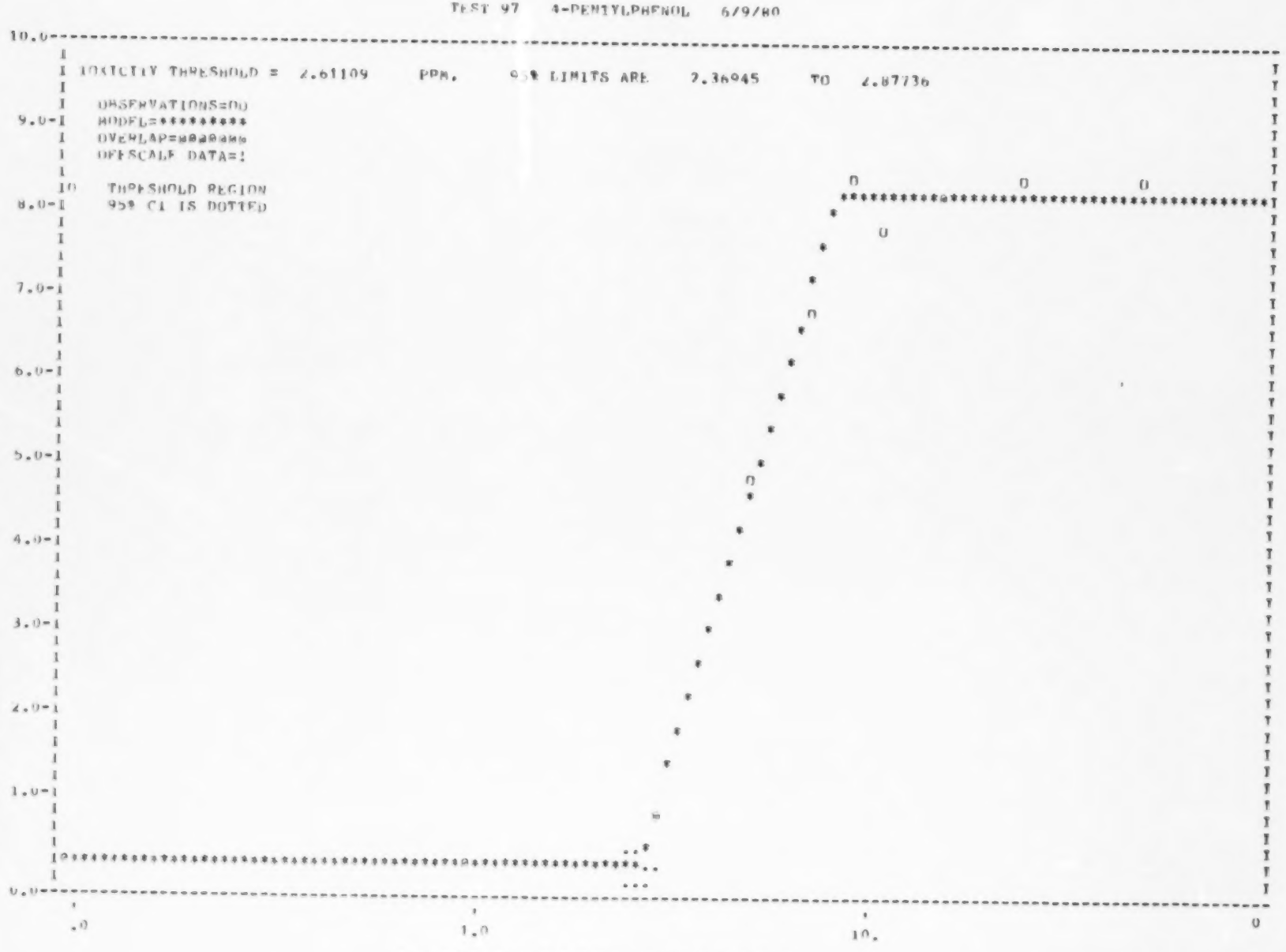
Figure 10(a) illustrates Version I where test data are entered directly through the terminal for a test with 4-pentylphenol. The threshold is computed to be 2.61 mg/l with the 95% confidence interval being 2.37 to 2.88 mg/l. Finally, the data are plotted at the terminal to show graphically how well the model fits the data (Figure 10(b)).

The chemicals tested in the evaluation of this screening method are presented in Table 4 and the correlation between the LC50 and the lethal threshold is presented in Figure 11. The correlation for this set of chemicals produced an $R^2 = 0.92$. Although the data indicate that a relationship exists between these two endpoints for toxicity, it is the current opinion that the use of the residual oxygen bioassay as a screening test is of marginal value. The test requires all of the fish culturing

Figure 10. Input of Residual Oxygen Test Data to Program HOCK

```
        IAS  PROGRAM DEVELOPMENT SYSTEM
               11:50:01          3-OCT-80

            ENVIRONMENTAL RESEARCH LAB SYSTEM


PDS> LOGIN EPARCR RCR
User EPARCR  UIC [350,40] TT01:  Job-id 333  11:50:13  3-OCT-80
NOTE:  NEW DEC-MOS MEMORY HAS BEEN INSTALLED (WED. 24-SEP-80).

LAST SYSTEM BACKUP:  TUES. 30-SEP-80    (4 1/2 TAPES WERE USED.) GETTING BETTER!

PDS> RUN HOCK
11:50:26
  PROGRAM HOCK(EY-STICK) : RESIDUAL DO BIOASSAY ANALYSIS

    WHAT IS THE INPUT FILE? IF TERMINAL TYPE      TI:
TI:
  ENTER PAIRWISE DATA ( [TOX] & R-DO ). CTRL-Z TO QUIT
0.09,0.39
0.96,0.24
2.9,0.96
4.8,4.91
6.8,6.99
8.7,8.48
10.6,7.84
14.5,8.22
24.,8.59
48.,8.48
97.,8.10
^Z


      #    [TOX]          LOG[TOX]          R-DO
      1  9.0000004E-02  -1.045758       0.3900000
      2  0.9600000      -1.7728778E-02   0.2400000
      3   2.900000       0.4623981       0.9600000
      4   4.800000       0.6812413        4.910000
      5   6.800000       0.8325090        6.990000
      6   8.700000       0.9395193        8.480000
      7   10.60000        1.025306        7.840000
      8   14.50000        1.161368        8.220000
      9   24.00000        1.380211        8.590000
     10   48.00000        1.681241        8.480000
     11   97.00000        1.986772        8.100000

  IS THAT OK? (YES OR NO)
YES
  NAME THE PLOT (120 CHARS. MAX.)
TEST 97   4-PENTYLPHENOL   6/9/80
  TYPE 5 TO GET THE OUTPUT HERE, TYPE 6 TO SEND PRINT IT
6
```

Figure 11. Residual Oxygen Test Results for 4-pentylphenol



Reproduced from best available copy.

Table 4

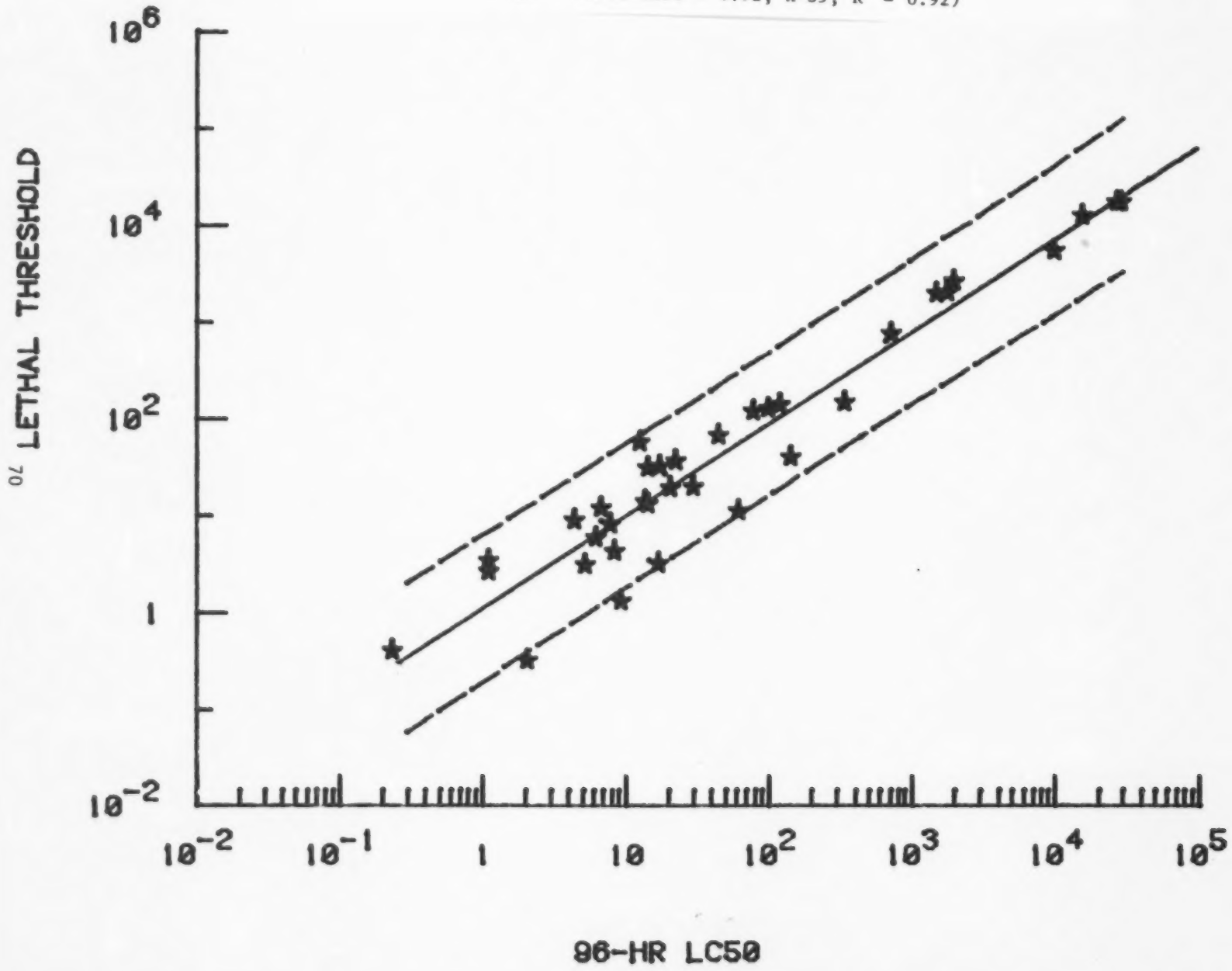
Compound	Lethal Threshold (mg/L)	96-Hour LC50 (mg/L)
Phenol	20	28.8
2,4-Dichlorophenol	4.2	8.23
2,4,6-Trichlorophenol	1.3	9.17
Pentachlorophenol	0.40	0.23
2,4,6-Tribromophenol	12	6.64
4-Nitrophenol	11	60.5
2,4-Dinitrophenol	3.2	16.7
2,6-Dimethylphenol	36	22.0
2,4-Dimethylphenol	32	16.8
2,4-Dinitro-o-crosol	0.32	2.04
3-Methoxyphenol	120	76.3
Tetrachloroethylene	14	13.5
1,1,2,2-Tetrachloroethane	19	20.3
2-Chlorophenol	58	12.4
1,1,2-Trichloroethylene	66	44.1
1,2-Dichloropropane	40	139
1,2-Dichloroethene	138	118

Table 4. (Continued)

Compound	Lethal Threshold (mg/L)	96-Hour LC50 (mg/L)
Methanol	17,600	28,200
Ethanol	12,700	15,000
2-Propanol	5,490	9,640
3,4-Xylenol	30.8	14
1-Butanol	2,090	1,740
sec-Butyl alcohol	2,650	1,910
1-Octanol	13	14.0
1-Hexanol	128	97.2
2-Methyl-1-propanol	2,020	1,460
1-Amino-2-propanol	147	327
o-Phenylphenol	5.8	6.0
1-Naphthol	8.9	4.24
4-Butylphenol	3.1	5.13
2-(2-Ethoxyethoxy) ethanol	17,700	26,363
4-Pentylphenol	2.6	1.09
Cyclohexanol	755	705
4-Phenylazophenol	3.4	1.09
4-Chloro-3-methylphenol	8.2	7.58

Figure 12. Relationship Between 96-hr LC50 and Lethal Threshold

(LT = 0.956 LC50 + 0.92, n=35, R^2 = 0.92)



and handling facilities of the 4 day acute test and, since the test requires an entire working day as a minimum, the time saved to produce an indirect estimate of the acute toxicity is judged to be insufficient to recommend its use as a screening test.

B. Microtox® Bacterial Bioluminesence Test

This rapid screening technique is the bacterial assay developed by Beckman Instruments Inc., which measures the decrease in natural light emission from the luminescent bacteria Photobacterium phosphoreum in response to a toxicant. The decrease in light is expressed as a 5-min EC50 (the concentration effecting a 50% reduction in light output).

Bacterial EC50s for 76 compounds were measured to evaluate the relationship between the EC50 and the 96-hr LC50 values for one species of fish, the fathead minnow (Pimephales promelas).

To eliminate variability in the fathead minnow LC50 values due to changes in water quality, chemicals were selected for bacterial testing only if the fathead minnow LC50s were derived from tests conducted at the Environmental Research Laboratory-Duluth. Chemical selection was further restricted to tests conducted in flow-through systems using Lake Superior water, and when exposure concentrations were measured.

Preparation of Test Solutions. All solutions for bacterial bioassay were prepared in distilled, deionized water. If the compound was readily soluble in water, concentrations were not measured and bacterial EC50 values were calculated from nominal concentrations. When testing less soluble compounds, a saturator consisting of a volumetric flask and a magnetic stirrer was used. Chemical concentrations for less soluble compounds were measured by gas chromatography.

Bacterial EC50 Determination. All bacterial bioassays were conducted using the Microtox Toxicity Analyser™ (Beckman Instruments, Inc., 6200 El Camino Real, Carlsbad, CA). Testing was done following the method described

by Beckman (1979) who supplied the reagents and lyophilized bacteria.

Inhibition of luminescence (γ) was calculated for each concentration:

γ = (initial light output x blank ratio/final light output) - 1

where the Blank ratio = final light output of the blank/initial light output of the blank.

A graph of $\log_{10}$ γ vs. $\log_{10}$ concentrations was plotted and the concentration causing 50% inhibition of luminescence was determined from that graph.

For most compounds it was possible to prepare a saturated stock solution that was at least double the EC50 concentration. When this was not possible, the saturated solution was tested and the reduction in light due to the toxicant was calculated as follows:

$$\frac{\text{mean chart reading for the Cl Blank}}{\text{mean chart reading for 100\% Toxicant solution}} - 1$$

Statistical Methods. The regression correlation for $\log_{10}$ 96-hour fathead minnow LC50 and $\log_{10}$ 5-minute bacterial EC50 was calculated for selected ethanes, alcohols, and ketones to determine the effects of chemical substitution on toxicity to fish and bacteria.

In addition, $\log_{10}$ EC50 vs. number of carbons and $\log_{10}$ fish LC50 vs. number of carbons was plotted for a homologous series of unsubstituted ketones and alcohols, and similar plots of EC50 and LC50 vs. number of chlorines were drawn for a homologous series of ethanes. Prediction intervals were calculated at the 95% level.

Evaluation of the Bacterial Bioassay. The correlations between EC50 and LC50 are summarized in Table 5. The correlation comparing 96-hour fathead minnow LC50s to 5-minute bacterial EC50s for unsubstituted alcohols (R^2=.96), unsubstituted ketones (R^2=.81) and chlorinated ethanes

(R^2=.98) were evident. Toxicity of chemicals to both fish and bacteria increased with addition of chlorine groups to ethanes. Addition of carbon groups also increased the toxicity of alcohols and ketones to bacteria and fish. A comparison of the reproducibility of acute fish and bacteria assay results (Table 6) shows no statistically significant differences in the standard deviation in the percent difference between replicates in the two tests.

The bacterial bioassay is a rapid and relatively inexpensive test. The entire test, including temperature equilibration for all reagents takes about 30 minutes to complete. The bacterial bioassay exhibits precision comparable to the precision of the acute bioassay with fish. Bulich (1979) found a coefficient of variance of 18.6 for 81 tests on sodium lauryl sulfate using a Microtox® Toxicity Analyzer.

From data obtained in this evaluation (Table 7), bacteria EC50 data could be used to screen certain types of chemicals for potential fish toxicity. This could be done using the prediction limits indicated in the graphs (Figure 12) and may be demonstrated using a hypothetical example in which a tier testing scheme included an acute toxicity trigger at an LC50 of 10 mg/L. If it were decided that extensive fish tests would not be required for chemicals with a LC50 greater than 10 mg/l, the data obtained in the evaluation could be partitioned accordingly. The point where the lower prediction limit intersects LC50 = 10 defines the EC50 (501 mg/l) corresponding to that fish toxicity value. The lower prediction limit is chosen to ensure that 95 percent of the compounds having an EC50 greater than the value will have an LC50 greater than 10 mg/l. Based on these data for organic chemicals, chemicals with an EC50 greater than 501 mg/l would have an LC50 greater than 10 mg/l and additional fish tests would not be necessary.

Table 5. Correlations for Classes of Chemicals of Log_{10} LC50 (96-Hr Fathead Minnow) vs. Log_{10} 5-Min EC50 (Bacteria)

Chemical Class	N	R^2
All Organic Chemicals	68	0.71
Unsubstituted Alcohols	9	0.96
Chlorinated Ethanes	4	0.98
Unsubstituted Ketones	10	0.81

Table 6. Summary of 5-minute EC50 and 96-Hr LC50 Results of Replicate Testing of Selected Chemicals in Bacteria and in Fathead Minnows

Chemical Compound	5-Min Bacteria EC50 (mg/1)	Percent Difference in Replicated EC50s
Acetone	22,000 21,000	5.0
2-Butanone	5,750 4,350	28.0
2-(2-Ethoxyethoxy) Ethanol	1,290 1,000	25.3
Hexanol	42.7 40.4	3.12
2-Methyl-2,4-Pentanediol	3,300 3,200 2,710	7.8
2-Octanone	20.5 15.0	30.98
Phenol	39.8 40.7	2.23

Chemical Compound	96-Hr Fathead Minnow LC50 (mg/1)	Percent Difference in Replicated LC50s
1-Amino-Propanol	327 287	15.5
Butanol	16.2 13.4	19.0
Cyclohexanone	732 527	32.56

Table 7. 96-Hr Fathead Minnow LC50s and 5-Min Bacterial EC50s.

Chemical Name	96-Hr Fathead Minnow LC50 (mg/1)	5-Min Bacterial EC50 (mg/1)
1-Decanol	2.3	1.16
*1-Octanol	13.4	6.3
1-Amino-2-Propanol	280	27.2
1-Amino-2-Propanol	327	27.2
2-Pehnoxyethanol	346	32.7
1-Hexanol	97.2	40.2
1-Hexanol	97.2	40.4
Cyclohexanol	705	115
2,3-Dibromopropanol	71	320
2-Methyl-1-Propanol	1,460	1,670
2,2,2-Trichloroethanol	298	1,800
1-Butanol	1,740	2,300
2-Methyl-2,4-Pentanediol	10,700	2,710
2-Methyl-2,4-Pentanediol	10,700	3,200
2-Methyl-2,4-Pentanediol	10,700	3,300
2-Chloroethanol	37	13,400
2-Propanol	9,640	35,000
Ethanol	14,200	44,000
Methanol	28,200	125,000

Table 7. (Continued)

Chemical Name	96-Hr Fathead Minnow LC50 (mg/l)	5-Min Bacterial EC50 (mg/l)
*2-Decanone	5.7	6.1
*2-Decanone	5.7	9.7
*2-Octanone	37	15
6-Methyl-5-Hepten-2-one	85.8	17.5
Cyclohexanone	527	18.5
Cyclohexanone	732	18.5
*2-Octanone	37	20
4,4-Dimethyl-Amino-3-Methyl-2-Butanone	8.4	42.1
4-Methyl-2-Pentanone	509	80
2-Butanone Oxime	844	950
5-Methyl-2-Hexanone	158	980
2,4-Pentanedione	96	1,050
5-Methyl-2-Hexanone	158	1,448
2-Butanone	3,200	4,350
2-Butanone	3,200	5,750
Acetone	8,140	21,000
Acetone	8,140	22,000
*p-tert-Butylphenol	5.15	0.21
Pentachlorophenol	0.220	0.08
4-Chloro-3-Methylphenol	7.59	0.58
Phenylazophenol	1.09	0.96

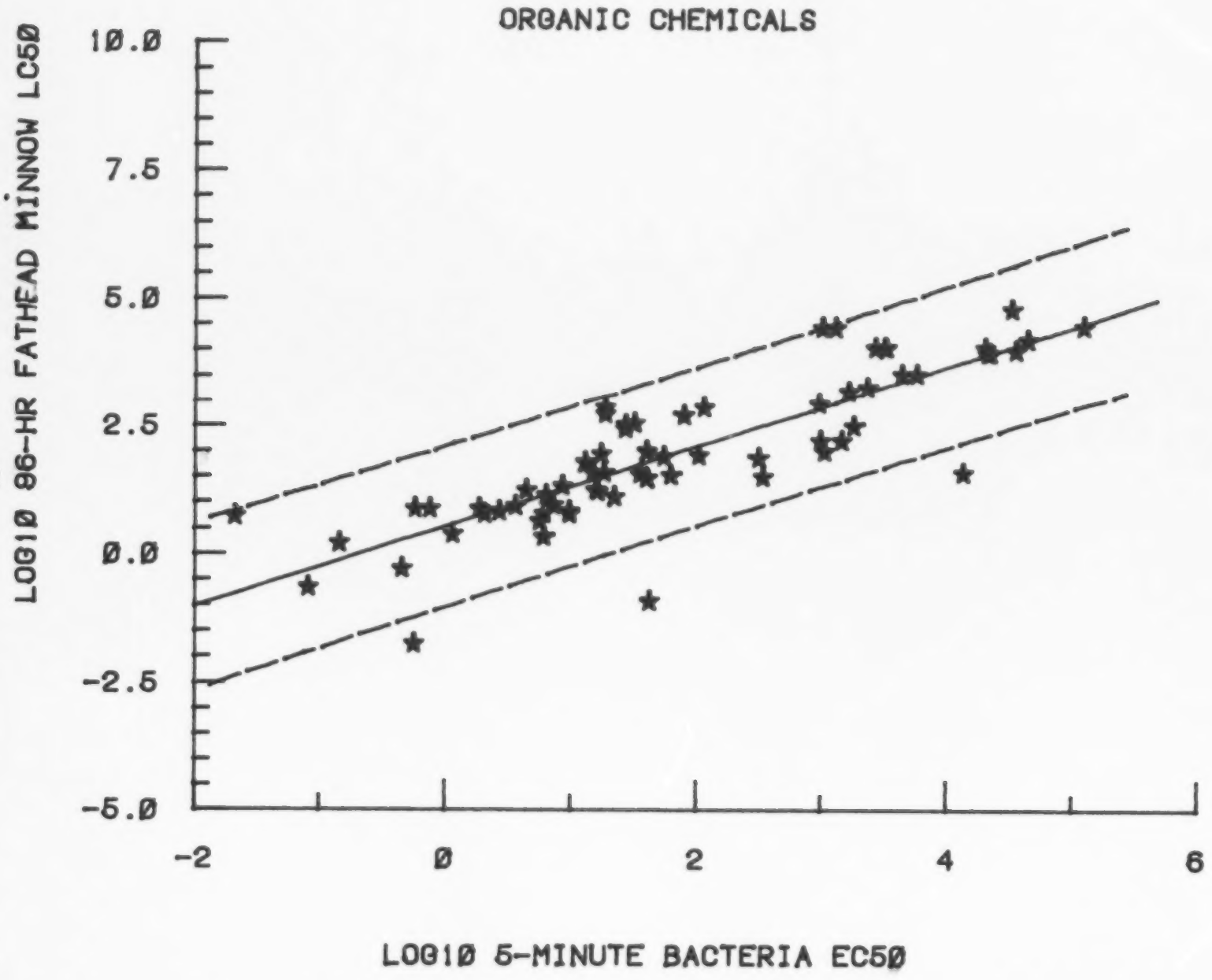
Table 7. (Continued)

Chemical Name	96-Hr Fathead Minnow LC50 (mg/l)	5-Min Bacterial EC50 (mg/l)
4-Chloro-3-Methylphenol	7.59	1.86
*O-Phenylphenol	6.0	2.05
2,4,6-Tribromophenol	6.64	2.7
2,4-Dichlorophenol	8.23	3.63
2,4-Dimethylphenol	16.81	4.4
*1-Napthol	4.24	5.66
4,6-Dinitro-o-Cresol	2.04	6.6
2,4,6-Trichlorophenol	9.17	7.20
*2-Allylphenol	15.9	10
*p-Nitrophenol	60.5	13.0
2,4-Dinitrophenol	16.7	15.8
o-Chlorophenol	12.4	22.1
*4-Amino-2-Nitrophenol	34.3	35.9
Phenol	28.8	39.9
Phenol	28.8	40.4
*Hexachloroethane	1.53	0.14
*Pentachloroethane	7.3	0.75
*Tetrachloroethane	20.3	8.6
*Trichloroethane	81.7	105
*Kelthane	0.51	0.45
*Permethrin	0.017	0.56

Table 7. (Continued)

Chemical Name	96-Hr Fathead Minnow LC50 (mg/l)	5-Min Bacterial EC50 (mg/l)
*Diazinon	6.65	9.8
Dimethyl Formamide	10,600	20,000
*N-Butyl Ether	32.5	63.0
2-(2-Ethoxyethoxy)-Ethanol	26,400	1,000
2-(2-Ethoxyethoxy)-Ethanol	26,400	1,290
Triethylene Glycol	61,000	33,000
*Butanal	16.2	16.4
*Ethanal	31.2	342

Figure 13. Relationship Between 96-hr LC50 and Microtox® EC50



If this screening technique had been applied to the compounds in this evaluation, fish tests would not have been performed on 35 of the 76 compounds. Because a bacterial bioassay costs considerably less than the estimated $2,000 for a fish test, the savings would be substantial.

Compounds which fall between the upper and lower prediction limits for LC50 = 10 would require additional testing with fish. Another decision scheme might use the screening test to circumvent intermediate testing on compounds likely to be highly toxic and "flag" them for long term testing. Those compounds could be determined from the regression plot using the same process described above and substituting the upper prediction limit. As testing continues on several species at the ERL-Duluth, it will be possible to make similar comparisons on other homologous series of chemicals. At present it appears that bacteria can be used to determine the relative toxicity to fish of selected groups of organic chemicals.

C. Algal Fluorescence Test

The "Algal Fluorescence Test" being developed is a sensitive and rapid test procedure based on the fluorescence of chlorophyll *a* from algae. Normally, chlorophyll *a* in algae will re-emit a small percentage of absorbed light as red fluorescence, the remainder being used in photosynthesis. If the overall health of the algal cell deteriorates the percentage of fluorescence increases. The increase in fluorescence has been determined to be inversely proportional to the photosynthetic capacity of the cell. By poisoning the cell with DCMU, a very potent herbicide, it is possible to totally block photosynthesis and achieve maximum fluorescence. DCMU causes a larger percent fluorescence increase in healthy cells, whereas in cells stressed by a toxicant this increase will be reduced. Using these principles we are refining our techniques for assaying toxicity using algae.

Cultivation. The green algae *Selenastrum capricornutum* and the blue green algae *Anabaena flos-aquae* were obtained from EPA-Corvallis, Oregon and were cultivated in AAP medium (U.S. EPA, 1978). Algal cultures were maintained on magnetic stir plates at 25°C. All cultures and test systems received continous irradiance of 350 ft. c. provided by 40 W "cool white" fluorescence bulbs. Exponentially growing cultures were used in all tests.

Test Procedures. The test system consisted of 24-5 ml test tubes in a plexiglass rack. Algae were either innoculated into lake water containing various concentrations of toxicant or the toxicant was added directly to an algae culture in AAP medium. Final test volume was 4 ml. Volatile compounds was tested in stoppered tubes. The test system was kept under continuous light until actual fluorometer readings were taken, usually 2 hours.

Fluorescence Measurements. All fluorescence measurements were done on a Baird-Atomic, Model SFR 100 Spectrofluorimeter at an excitation of 430 nm and an emission of 680 nm. The test system was kept under continous illumination until actual measurements were taken so as not to alter the photosynthetic state of the culture. After the incubation period the culture tube was read directly in the fluorometer and an initial reading was recorded after the signal stabilized (F_N). Similarly a second reading was recorded after the addition of enough DCMU (3-(3,4-dichlorophenyl)-1,1-dimethyl-urea), a herbicide which blocks the electron transport pathway in photosynthesis, to obtain a 5 μM solution (F_D). These readings were used to calculate a percent fluorescence increase:

$$\frac{(F_D - F_N)}{F_N} \times 100 = \% \text{ fluorescence increase.}$$

These values were compared statistically to the control to detrmine effect concentrations of a variety of organic compounds.

Evaluation of the Algal Fluorescence Test. Typical results with several organic toxicants are presented in Table 8. Although data are insufficient to be statistically analyzed with data from other algal tests, the results indicate the relative sensitivity of algae to selected compounds. A dose-response curve is presented in Figure 14. The method appears to be sensitive, reliable, and fast (2 hours) to a broad range of toxicants. With the chemicals tested thus far, effect concentrations are generally in the range of the fathead acute values being generated. The initial results with natural surface waters being tested for the "Ambient Metals Study - EPA" indicate that the method can also detect metal toxicity.

The method will continue to be refined with respect to test techniques for analyzing the data. A systematic testing of those chemicals being used

Table 8. Effect concentrations using the algal fluorescence test

Compound	Algae	Effect[a] Concentration	Test Duration
Ethanol	Selenastrum	50 ppm*[b]	2 hr
Methanol	Selenastrum	>32,000 ppm	2 hr
Butanol	Selenastrum	300 ppm**	2 hr
Hexanol	Selenastrum	>320 ppm	2 hr
Hexanal	Selenastrum	80 ppm*	2 hr
Hexanal	Anabaena	160 ppm*	2 hr
Acetone	Selenastrum	4,000 ppm*	2 hr
DMF	Selenastrum	6,250 ppm**[b]	2 hr
Tetrachlorophenol	Selenastrum	15.4 ppm**	2 hr
1,2,4-Trichlorobenzene	Selenastrum	8 ppm*	24 hr
1,3-Dichlorobenzene	Selenastrum	60 ppm**	3 hr
1,1,2,2-Tetrachloroethane	Selenastrum	51.1 ppm**	1.5 hr

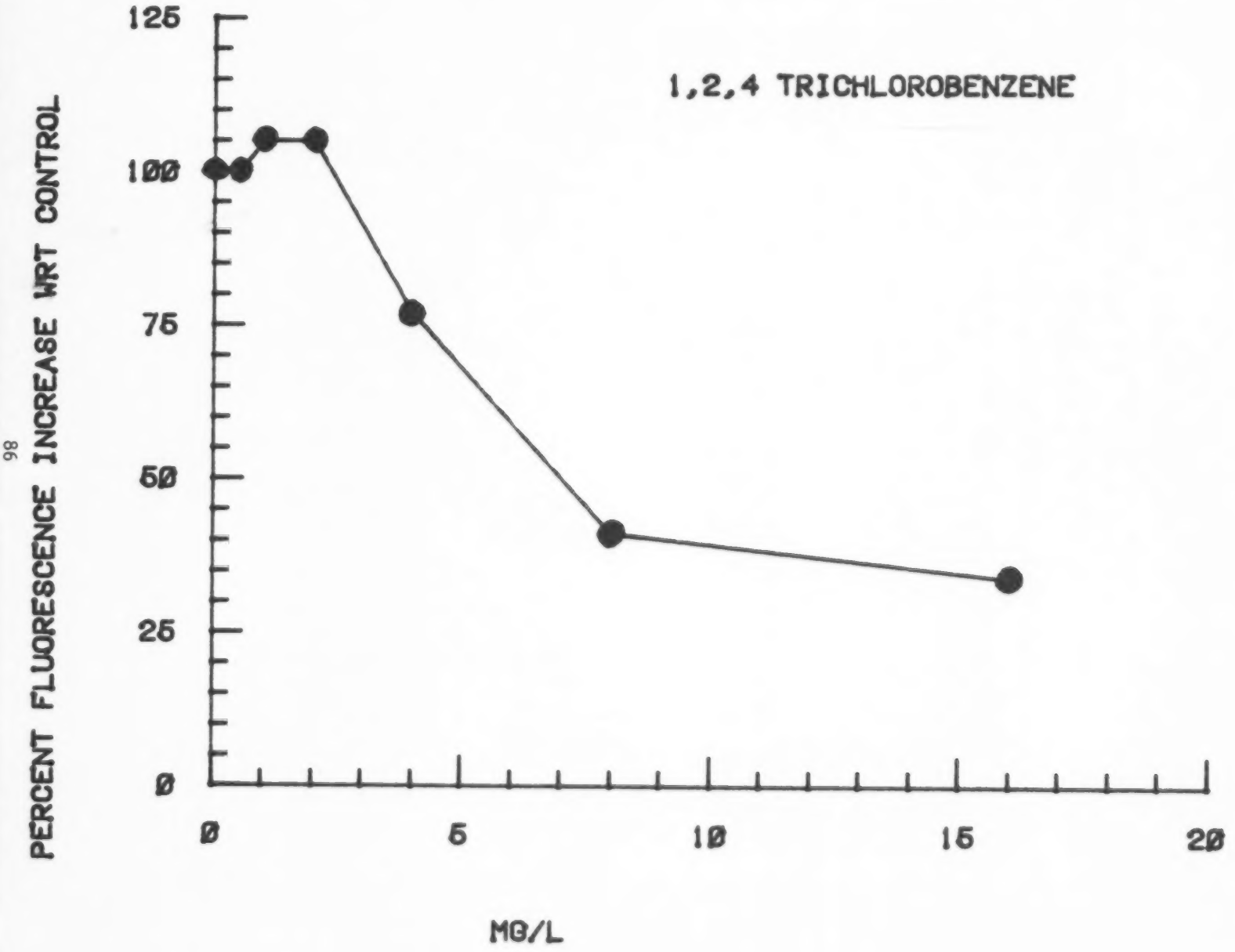
[a] Lowest concentration tested showed significant difference from control

[b] Lowest concentration tested

* P = 95

**P = 99

Figure 14. Dose-Response of Algal Fluorescence to Trichlorobenzene



in the QSAR program will be performed. Plans have been made to use a green algae and a blue green algae to determine comparative sensitivities to test compounds.

The literature and data generated thus far suggest that the reduction or enhancement of chlorophyll *a* fluorescence before the addition of DCMU may not only indicate toxicity, but also the site of action of the particular inhibitor. This variable fluorescence response has the potential to delineate between various modes of action of structurally different compounds. For example, DCMU type compounds block non-cyclic photosynthetic electron transport and increase F_N, whereas benzoquinone, a well known Hill oxidant, lowers F_N. Along with our continued effort to improve the method we will also be considering these "behavioral" responses of chlorophyll *a* fluorescence in an attempt to describe mode specific chemical activity. A literature review on the potency and modes of action of chemicals in inhibiting the Hill reaction in photosynthesis will be the subject of subsequent reports.

X. LITERATURE REVIEW ON NARCOSIS

In an effort to summarize the major modes of action of common industrial chemicals, a series of literature reviews are planned. These reviews are an integral part of QSAR because they help to determine truely anomalous behavior within a model from changes in mode of action which suggest an additional model is needed. The first review is concerned with narcosis for the reasons stated previously in this report.

Narcosis is a reversible state of arrested activity of various protoplasmic structures under the influence of certain chemicals. Kaufman (1977) defines "general anesthesia as a pharmacologically induced reversible disruption resulting in a coordinated sequence of changes in neural activity". The terms narcosis and anesthesia are used interchangeably in medical literature. Narcosis of the central nervous system is referred to as general anesthesia. The chemicals or drugs producing arrested activity of any cellular organization are referred to as narcotics, biological depressants, hypnotics, and general anesthetics (Albert 1965). A variety of chemical agents of different chemical structures and properties produce narcosis in animals. These agents may be gases (Hesser et al., 1978) aliphatic or aromatic hydrocarbons, chlorinated hydrocarbons, alcohols, ethers, ketones, aldehydes, weak acids, weak bases, and some aliphatic nitrocompounds (Albert, 1965; Gero, 1965; Roth, 1980).

Most of the biological cells are susceptible to the actions of general anesthetics (Roth, 1980). However, it appears that all anesthetics or narcotics produce a general depression of the central nervous system (CNS) of the organisms. Vandam (1965) pointed out that whatever may be the mechanism of narcosis, the clinical signs of general anesthesia are related to the

amount of anesthetic made available to the CNS. Anesthetics may affect not only the CNS but a variety of cell types and several synthetic model preparations (Halsey et al., 1974; Miller and Miller, 1975; Roth, 1979; Seeman, 1972). In a recent study, Schwab and Schwab (1979) have shown that protozoan (Allogromia laticollaris Arnold) cells after treatment with 1 mM of the volatile anesthetic, halothane, showed disappearance of cytoplasmic microtubles. These workers also demonstrated the appearance of tubulin polymers which were helices of paracrystalline structures.

During the last few years new information has accumulated concerning the pharmacological effects of narcotic agents. With the development of novel monitoring devices and in vitro techniques, it has been demonstrated that anesthetics can exert profound and direct effects on cardiac muscles and alter regional blood flows in a number of organs (Price and Ohnishi, 1980; Longnecker and Harris, 1980). Many narcotic agents are capable of altering tissue and plasma concentrations of cyclic nucleotides such as cAMP or GMP and of neurohumoral substances such as catecholamines, dopamine, serotonin, and renin (Altura, 1980; Rahwan, 1975; Zaleska, 1979). In addition these anesthetics have the ability to perturb lipid-bilayer membranes and alter ion transport (Roth, 1980; Lee, 1976; Schettini and Furniss, 1979).

A mechanism of action of some reactive anesthetics has been sought in an inhibition of enzymic processes. For example, a number of anesthetics inhibit mitochondrial oxygen consumption. The mode of action in mitochondria appears to be chiefly inhibition of electron transport with slight uncoupling of oxidative phosphorylation (Altura, 1980).

In spite of intensive pharmacodynamics, physicochemical, biochemical and clinical research, the exact changes in the CNS remain unknown. It is quite puzzling how some chemically inert agents and others which are rather

reactive chemicals may induce reversible changes in the functions of the CNS in such a way that loss of consciousness and insensitivity to pain occurs (Kuschinsky and Lullmann, 1973; Adriani and Naraghi, 1977).

There are several hypotheses which have been proposed to explain the mode of action of narcosis. It is generally accepted that the majority of general anesthetics affect most biological cells in a non-specific fashion. However, due to their diverse chemical structures, a similar or unitary mechanism of action appears to be inconceivable. Recent studies suggest that all anesthetics may not act via the same mechanism (Stockard and Bickford, 1975; Trudell, 1977).

One of the oldest and most widely quoted theories of anesthetic action is the Meyer-Overton "lipoid hypothesis of narcosis". Meyer (1899) and Overton (1901) independently proposed that anesthetic action correlates with a high oil/water partition coefficient. K. H. Meyer (1937) stated that "narcosis (anesthesia) commences when any chemical substance has attained a certain molar concentration in the lipids of the cells." It is true that most anesthetics, especially volatile anesthetics, have high oil/water partition coefficients, but this does little to provide an explanation of how they function as depressants (Roth, 1980; Gero, 1965; Koblin and Eger, 1979). This theory may only explain why such anesthetics accumulate in the lipid rich CNS (Roth and Seeman, 1971). In addition, there are some outstanding exceptions to the Meyer-Overton hypothesis of anesthesia, such as alcohols which are nonspecifically acting anesthetics despite their low oil/water distribution coefficient (Albert 1965). Furthermore, oils and certain other substances with very high partition coefficients, do not produce narcosis.

Traube (1904) proposed a different physicochemical hypothesis of narcosis. This theory describes a lowering of surface tension between water

and a second phase such as oil or air. The alcohols in homologous series fit this generalization, but even familiar hydrocarbons like chloroform and ethylchloride do not lower interfacial tension at the surface boundary of water/air or water/oil. Furthermore, many detergents which are known to lower surface tension are not depressants.

Ferguson (1939, 1951) provided a physicochemical explanation based on the thermodynamic or chemical potential of the depressants in the internal phase and the biophase (site of biological action). He pointed out that a stable level of narcosis is achieved when an equilibrium of drugs is reached between the two phases. Ferguson's principle provides a satisfactory correlation between chemical potential and the activity of nonspecifically acting drugs, but does not explain the true mode of action of depressants. Depression of physiological function generally appears to be the typical action of compounds which do not enter into specific reactions with the organisms (Albert, 1965).

Pauling (1961) postulated that hydrophobic depressants cause the formation of hydrate microcrystals in the brain. Such hydrates are known to be formed at low temperatures around nonpolar molecules and nonpolar groups within molecules, and are referred to as "clathrates". These clathrates are ordinarily unstable at body temperature, but it is also known that clathrates formed around several kinds of molecules or groups (mixed clathrate hydrates) are more stable (Gero, 1965). Pauling further assumed that clathrate hydrates are formed in brain cells in the presence of volatile anesthetics which then combine with the alkyl groups of the proteins and phospholipids. These water clathrates are stable under physiological conditions and they cause water molecules in their vicinity to change to the denser structure of ice. The ice microcrystals formed in this way tie up water, increasing the

impedance of the neural network of conductors and hindering the movement of ions in the brain. Unlike other theories, Pauling's hypothesis centers around the aqueous phase and not on the lipid material of the CNS, and also seeks to connect depressant action with a process of reversible disorganization.

More recent theories of narcosis seek the mechanism of depressant action in interactions of anesthetics with cell membranes (McElroy, 1947; Miller and Pang, 1976; Rosenberg et al., 1975; Ashcroft et al., 1977ab; Haydon et al., 1977; Jain et al., 1978). It is well accepted that conduction of impulses in the neurons and muscle fibers depends on their cable structure and occurence of specific changes in the ion permeability of the surface membranes (Keynes, 1972). Biological membranes are composed of a phospholipid bilayer and cholesterol, arranged with polar (hydrophillic) headgroups facing outward and the lipophillic or hydrophobic moeity inward. The protein of the membrane on either side of the phospholipid bilayer forms a trilaminar sandwich (Keynes, 1972). Membrane structures have pores or channels through which sodium and potassium can move across the membrane during the transmission of nerve impulses (Ashcroft et al., 1977). Most investigators have envisioned membrane lipids as the target of anesthetic actions. However, some believe that general anesthetics inactivate proteins, which are essential for CNS function (Eyring et al., 1973; Kaufman, 1977).

Recent Theories of Anesthesia

(a) Phase Transition - Three recent theories propose that anesthetics disrupt the lipid bilayer of the membrane. Lee (1976) postulated an "annular transition model", also called a "phase transition hypothesis". This hypothesis proposed that the sodium channel of an excitable membrane is surrounded by an annulus of lipid which is in the "crystalline" or gel state

(solid phase), which helps to maintain an open sodium channel. Addition of an anesthetic triggers a change in the lipid to a fluid or liquid crystalline state and this alteration allows the closure of the sodium channel, resulting in anesthesia. An increase in temperature can also alter the gel phase to a liquid phase, and the temperature at which this transition occurs is called the phase-transition temperature (Roth, 1979). The "phase transition hypothesis" is supported by the studies of Hill (1978) on a phospholipid membrane model, and of Kamaya et al. (1979) who showed that high pressure antagonizes both anesthesia and the anesthetic-induced decrease in phase transition temperature of dipalmitoyl phosphatidylcholine bilayers. However, a recent study by Pringle and Miller (1978) which utilized structural isomers of tetradecanol, indicated no direct relationship between lipid phase transition and the mechanism of anesthesia. Gases, barbiturates, steroids and alcohol anesthetics have all been shown to fluidize phosphatidylcholine-cholestrol lipid bilayers and some biological membranes. Miller and Pang (1976) using different lipid composition and anesthetics demonstrated that the fluidizing efficacy of an anesthetic varied with different anesthetics and also with the lipid composition of the membrane.

(b) Lateral Phase Separation - This hypothesis by Trudell (1977) proposes that lipids under normal conditions exist as gel and liquid phases. The region where a gel phase coexists with a liquid phase is defined as a lateral phase separation (Trudell, 1977). Conversion of one phase to another phase allows the membranes to expand or contract with less energy than would be necessary if the membrane was only in gel or fluid form. Also the regions (fluid or gel) may be associated with functional proteins. Trudell proposed that anesthetics may reduce the gel phase in membranes and thus decrease the lateral phase separation. This decrease in the lateral phase separation

would prevent the functional protein from altering its conformation necessary for translocation of ions across the post synaptic membrane or for the release of a neurotransmiter, thus resulting in depression or narcosis (Trudell, 1977; Trudell et al., 1975).

(c) Fluidized Lipid Hypothesis - This hypothesis postulates that anesthetics can disorder the motion of the lipid membrane components (Gage and Hamill, 1976), and that due to increases in the lateral and rotational motion of the lipid component, the stability of the channel proteins is reduced. This facilitates the closure of the ionic channels in the membrane (Gage and Hamill, 1976; Miller and Pang, 1976; Miller, 1977; Roth, 1980).

Theories discussed above as "phase transition" (Lee 1976), "lateral phase separations" (Trudell 1977) and the "fluidized lipid" hypothesis all propose that the phenomenon of narcosis is dependent on fluidization of the lipid membrane. The spin label study of Pringle and Miller (1978) showed that unsaturated long chain alcohols and cis or trans-tetradecanol could fluidize egg lecithin cholesterol membranes to similar levels. However, contrary to that, cis-tetradecanol lowered the transition temperature of dipalmitoylphosphatidylcholine while trans-tetradecanol elevated this transition temperature. These results are consistent with a lipid fluidity model of anesthetic action, but provide little support for a "lateral phase separation" theory of anesthesia. The authors suggest direct interaction of alcohols with the membrane protein (Pringle and Miller, 1978). Studies conducted by Jain et al. (1975) on the transition profile of dipalmetoyl lecithin liposomes exposed to a variety of drugs support the theory of fluidity disorder of the lipid bilayer.

(d) Critical Volume Hypothesis - Mullins (1954) proposed a critical volume hypothesis. According to this hypothesis, narcosis is produced by

expansion of the critical hydrophobic site due to the adsorption of anesthetic molecules. Mullins (1954) suggested that anesthetic molecules should have a complementary shape or size to "fit" into the free spaces of the membrane's chemical lattice. This critical volume of occupation will thus result in decreased sodium permeability and depression of membrane excitability.

Several studies lend support to Mullin's critical volume expansion hypothesis. For example, application of hydrostatic pressure (pressure reversal of anesthesia) removed, at least partially, the effects of general anesthetics in certain amphibians and mice (Lever et al., 1971; Miller et al., 1973). The phenomenon of pressure reversal of anesthesia was first observed in luminous bacteria (Johnson et al., 1942), which led Lever et al. (1971) to formulate a unifying concept of critical volume expansion as a general mechanism of narcosis. They postulated that anesthetics expand the size of a critical hydrophobic site, while the hydrostatic pressure contracts this site and causes reversal of anesthesia. Lever et al. (1971) calculated that general anesthetics at critical concentrations in the lipids would produce an expansion of lipids of the order of 0.4%. Further studies on quantitative measurements of erythrocyte membranes (Roth and Seeman, 1971, 1972ab), and of lipoprotein films (Clements and Wilson, 1962) support the estimate of 0.5% expansion of biomembranes in the presence of general anesthetics. High pressure of approximately 100 atmospheres indeed antagonizes the effect of general anesthesia, but the degree of the high pressure reversal varies with different anesthetics, and the antagonism of high pressure may not be proportional or linear with the pressure (Halsey et al., 1978).

A multisite expansion hypothesis was proposed by Halsey et al. (1978). They stated that a general anesthetic may influence the expansion of more than one molecular site with different physical properties, and the physical properties of molecular sites may be influenced by anesthetics and pressure. In addition, the molecular sites have a finite size and limited occupancy, and the pressure does not necessarily influence the same site as anesthetics. Finally the molecular sites for narcosis may not be disturbed by decreases in temperature in a manner analogous to increases in pressure. The multisite expansion hypothesis based on intact animal studies is more relevant, particularly with complete membranes in which the molecular sites are energetically linked together (Halsey et al., 1978).

(e) Degenerate Protein Perturbation Hypothesis - Previous studies discussed in this review all support a common structural perturbation in the excitable membrane which is caused by anesthetics of diverse chemical structures. This unitary hypothesis of narcosis was derived from experimental evidence that anesthetic potency correlates approximately with oil/water partition coefficients of narcotic agents (Seeman, 1972; Miller et al., 1977). However, direct experimental evidence for any unitary mechanism has not been presented to describe the interactions of anesthetics with target protein structures (Richards et al., 1978). Studies conducted by Richards et al. (1978) utilizing n-alkanols do not provide support to a unitary mechanism of narcosis. These workers have suggested that anesthesia is produced by direct interactions between the anesthetic molecules and target membrane protein. They assume that smaller anesthetic molecules (halothane or cyclopropene) distribute in one set of hydrophobic sites of appropriate dimensions within the protein, while larger molecules like barbituates, steroids, and aromatic amines may bind to distinct sets of sites

on the target protein (Richards et al., 1978). Eyring et al. (1973) had also proposed that anesthetics may interact with the hydrophobic region of a protein and change its conformation to make it less active. The protein pertubation theory of anesthesia does not provide a distinct mechanism of narcosis (Koblin and Eger, 1979), rather indirect evidence is derived from thermodynamic principles (Eyring et al., 1973) and the lack of correlation between lipid membrane fluidization with n-alkanols.

(f) Cyclic AMP Metabolism - Cohn and Cohn (1980) proposed that anesthetics interact with the activity of a specific enzyme protein hydrophobic site which binds with the adenine moeity of cAMP. According to these authors cAMP *in vivo* regulates the duration of narcosis without altering the cAMP concentrations in the brain. These workers have shown that pentobarbital causes dose dependent inhibition of brain adenylate deaminase which converts AMP to inosine (Cohn and Cohn, 1980). It has been shown that dibutyryl cyclic AMP shortens the duration of narcosis induced by a wide variety of narcotics and hypnotics, and protects against the lethal doses of amobarbital (Cohn et al., 1980). The effects of anesthetics on transmembrane regulatory systems may require further studies. The role of hormone and neurotransmitter-dependent adenylate cyclase is known in transmembrane regulatory systems and production of cAMP (Rimon et al., 1978).

Our knowledge of the mode of action of anesthetic chemicals awaits further understanding of the structure of phospholipid bilayers, axonal transport, conformation of nerve proteins, synaptic transmission in the central nervous system and the interactions of anesthetics with endogenous components of the neurons. Current theories of narcosis emphasize lipid membrane disorders in the CNS without specificity (Roth, 1980).

In mammalian systems, four definite stages of anesthesia can be distinguished after the administration of anesthetics (Vandam, 1965). Stage I anesthesia is characterized by motor excitation and inability to coordinate motor activity (ataxia). Stage II anesthesia is characterized by delirium when the animals show hallucinatory and cataleptic behaviour. Stage III anesthesia, also called surgical anesthesia, is further progression of stage II with increasing immobility and loss of reflexes. During this stage the subject becomes unresponsive to painful stimuli. Further increases in anesthetic dosage may cause a depression of the total central nervous system activity (Stage IV), and irreversible damage to the brian which results in medullary paralysis and death. These different stages of anesthesia classified by Guedel (1951) as unidirectional schema were revised on the basis of neurophysiological studies (Winters, 1976).

Fishes treated with anesthetic agents have been shown to undergo similar stages of anesthesia (McFarland and Klontz, 1969; Klontz and Smith, 1968; Bonath, 1977). The phenomenon of narcosis in fishes has not been extensively studied. Hunn and Allen, (1974) reviewed the recent studies on MS222 and quinaldine, used widely as fish anesthetics. Such compounds are readily absorbed through the gills of fish, as demonstrated with rainbow trout (Salmo gairdneri), channel catfish (Ictalurus punctatus), marine dog fish (Squalus acanthias), brook trout (Salvelinus fontinalis), bluegills (Lepomis macrochirus) and largemouth bass (Micropterus salmoides). High lipid solubility and rapid rates of diffusion across the gill membrane appear to be important properties for fish anesthetic agents (Hunn and Allen, 1974). For example, MS222 has been shown to be readily taken up through the gills of freshwater fishes according to its concentration gradient. From the gills it is distributed to many body tissues, especially the cerebrospinal fluid (CSF)

in the brain (Hunn, 1970; Hargens et al., 1974; Stenger and Maren, 1974; Daxboeck and Holeton, 1980).

Fish have been shown to possess superficial and internal (or central) oxygen receptors which initiate active cardiac and ventilatory responses under environmental hypoxia. Anesthetics may interact with these receptors (Daxboeck and Holeton, 1978; Smith and Jones, 1978; Bamford 1974a,b; Marvin and Burton, 1973). Mechanisms by which such chemical agents disrupt various physiological functions in fishes largely remains unknown and deserves further studies.

Studies conducted in our laboratory with the fathead minnow (Pimephales promelas) to determine 24, 48, and 96 hour median lethal concentrations (LC50) of a group of alcohols, aldehydes, ketones and ethers indicate that these chemicals cause fish immobility, hyperexcitation and loss of equilibrium before the onset of toxic symptoms. These symptoms are apparently dependent on the chemical dosage and the exposure time. It is probable that these chemicals produce lethal effects due to medullary paralysis, as shown in mammals after dosing with higher concentrations of the narcotics. The typical signs of the stage I anesthesia (indicated by motor excitation and ataxia) in fish may be determined in vivo without killing the test fish and it may provide useful insight.

Previous studies on long chain alcohols have shown that they induce narcosis and anesthesia in different animal species. Jain and Wray (1978) have shown that the lipid bilayer/water partition coefficient for n-alkanols are at least 3 fold smaller than their partition coefficient in the bulk solvents. They imply that alcohols may induce changes in the lipid biomembranes at much lower concentrations than can be predicted from their octanol/H_2O partition coefficients (Jain and Wray, 1978). Similar studies

on the partition coefficients of a variety of chemicals with actual biological tissues (e.g. with fish gill/H_2O or blood/H_2O) may provide useful informtion on the movement and uptake of these chemicals, and may have some degree of correlation with their toxicity. McCreery and Hunt, (1978) studied 62 compounds, including a broad range of alcohols for their abilities to intoxicate rats in vivo. These workers suggested that the shape of a molecule had little effect on its ability to intoxicate, but the compound's amphiphillicity was important for its ability to partition into neuronal membranes (McCreery and Hunt, 1978). Several studies have indicated that membrane/water partition coefficients are more meaningful than partition coefficients with artificial solvents (Roth and Seeman, 1972a,b; Jain and Wray, 1978; Mullins, 1954; Seeman, 1972). Biologically active compounds must reach a definite target site in order to elicit a particular response, and such compounds cross multiple hydrophyllic and hydrophobic barriers before reaching the site of action (Kubinyi, 1978).

REFERENCES

Adriani, J. and M. Naraghi. 1977. The pharmacologic principles of regional pain relief. Ann. Rev. Pharmacol. 17: 223-242.

Albano, C., W. Dunn III, U. Edlund, E. Johansson, B. Norden, M. Sojostrom and S. Worl. 1978. Four levels of pattern recognition. Anal. Chim. Acta 103: 429-443.

Albert, A. 1965. Selective Toxicity. 3rd edition, John Wiley & Sons, Inc., New York, pp. 3-380.

Altura, B. M. 1980. Cardiovascular actions of general anesthetics. Fed. Proceed. 39: 1574.

Ashcroft, R. G., H. G. L. Coster, and J. R. Smith. 1977a. Local anesthetic benzyl alcohol increases membrane thickness. Nature 269: 819-820.

Ashcroft, R. G., H. G. L. Coster, and J. R. Smith. 1977b. The molecular organization of bimolecular lipid membranes. Biochem. Biophys. Acta 469: 13-22.

Bamford, O. S. 1974a. Oxygen reception in the rainbow trout (Salmo gairdneri. Comp. Biochem. Physiol. 48A: 69-76.

Bamford, O. S. 1974b. Respiratory neurons in rainbow trout (Salmo gairdneri). Comp. Biochem. Physiol. 48A: 77-83.

Biagi, G. L., A. M. Barbaro, M. C. Guerra, M. Babbini, M. Gaiardi, and M. Bartoletti. 1980. R_m values and structure-activity of benzodiazepines. J. Med. Chem. 23: 193-201.

Bonath, K. 1977. Information on experimental animals ±4 Narcosis of repitiles, amphibians and fish, Part 3. pp. 99-149. Published by Verlag-Paul Parey, Berlin, Germany.

Bulich, A. A.,M. W. Greene, and D. L. Isenberg. 1979. The reliablity of one bacterial luminescence assay for the determination of toxicity of pure compounds and complex effluents. ASTM, Proceedings of Chicago Meeting, 1979, (unpublished).

Cammarata, A. 1972. Interrelationship of the regression models used for structure-activity analyses. J. Med. Chem. 15: 573-577.

Cammarata, A. and G. K. Menn. 1976. Pattern recognition classification of therapeutic agents according to pharmacophores. J. Med. Chem. 19: 739-747.

Chu, K. C., R. J. Feldmann, M. B. Shapiro, G. F. Hazard, Jr., and R. I. Geran. 1975. Pattern recognition and structure-activity relationship studies. Computer-assisted prediction of antitumor activity in strkucturally diverse drugs in an experimental mouse brain tumor system. J. Med. Chem. 18(6): 539-545.

Clements, J. A. and K. M. Wilson. 1962. The affinity of narcotic agents for interfacial films. Proceed. Natl. Acad. Sci. USA 48: 1008-1014.

Cohn, M., M. L. Cohn, S. K. Wolfson, and F. M. Steichen. 1980. Dibutyryl cyclic AMP - an effective antidote against lethal doses of amobarbital in rhesus monkey. Fed. Proceed. 39: 406.

Cohn, M. L. and M. Cohn. 1980. A new molecular theory of anesthesia. Fed. Proceed. 39: 406.

Craig, P. N. 1971. Interdependence between physical parameters and selection of substituent groups for correlation studies. J. Med. Chem.14: 680-684.

Daxboeck, C. and G. F. Holeton. 1978. Oxygen receptors in the rainbow trout (Salmo gairdneri). Can .J. Zool. 56: 1254-1259.

Daxboeck, C. and G. F. Holeton. 1980. The effect of MS222 on the hypoxic response of rainbow trout (Salmo gairdneri).

Draper, N. R. and H. Smith. 1966. Applied Regression Analysis. Wiley and Sons, New York, NY.

Dunn, N. J., III, and S. Wold. 1978. A structure-carcinogenicity study of 4-nitroquinoline 1-oxide using the SIMCA method of pattern recognition. J. Med. Chem. 21: 1001-1007.

Eyring, H., J. W. Woodbury, and J. S. D'Arrigo. 1973. A molecular mechanism of general anesthesia. Anesthesiology 38: 415-424.

Ferguson, J. 1939. The use of chemial potentials as indices of toxicity. Proceed. Roy Soc. London 127B: 387-404.

Ferguson, J. 1951. Mechanisme de la narcose. IN: Colloques Internationaux du Centre National de la Recherche Scientifique, Paris. p. 25.

Frank, R. E., W. E. Massy and D. G. Morrison. 1965. Bias in multiple discriminant analysis. J. Market. Res. 2: 250-257.

Fujita, T. and T. Ban. 1971. Structure-activity of phenethylamines as substrates of biosynthetic enzymes of sympathetic transmitters. J. Med. Chem. 14: 148-152.

Gage, P. W. and P. O. Hamill. 1976. Effects of several inhalation anesthetics on the kinetics of postsynaptic conductance changes in mouse diaphram. Br. J. Pharmacol. 57: 263-272.

Gero, A. 1965. Intimate study of drug-action III: Possible mechanism of drug action IN: Drill's Pharmacology in Medicine (ed. J. R. DiPalma), Blackiston Div., McGraw Hill Book Co., New York, 3rd edition, pp. 47-69.

Glave, W. R. and C. Hansch. 1972. Relationship between lipophilic character and anesthetic activity. J. Pharm. Sci. 61(4): 589-591.

Gray, N. A. B. 1976. Constraints on "learning machine" classification methods. Anal. Chem. 48: 2265-2268.

Guedel, A. E. 1951. Inhalation anesthesia, 2nd ed. pp. 10-52. McMillan Co., New York.

Halsey, M. J., R. A. Millar, and J. A. Sutton. 1974. Molecular mechanisms in general anesthesia. Published by Churchill-Livingstone, Edinburgh, England.

Halsey, M. J., B. W. Smith, and C. J Green. 1978. Pressure reversal of general anesthesia - a multisite expansion hypothesis. Br. J. Anaesth. 50: 1091-1097.

Hansch, C. and T. Fujita. 1964. p-σ-π analysis A method for the correlation of biological activity and chemical structure. J. Amer. Chem. Soc. 68: 1616-

Hansch, C., J. E. Quinlan, and G. L. Lawrence. 1968. The linear free-energy relationship between partition coefficients and the aqueous solubility of organic liquids. J. Org. Chem. 33(1): 347-350.

Hansch, C., A. Leo, S. H. Unger, K. H. Kim, D. Nikaitani, and E. J. Lien. 1973. "Aromatic" substituent constants for structure-activity correlations. J. Med. Chem. 16(11): 1207-

Hansch, C, S. H. Unger, and A. B. Forsythe. 1973. Strategy in drug design. cluster analysis as an aid in the selection of substituents. J. Med. Chem. 16: 1217-1222.

Hargens, A. R., R. W. Millard, and K. Johansen. 1974. High capillary permiability in fishes. Comp. Biochem. Physiol. 48A: 675-680.

Haydon, D. A., B. M. Hendry, S. R. Levinson, and J. Requena. 1977. Anesthesia by n-alkanes - A comparative study of nerve impulse blockage

and the properties of black lipid bilayer membranes. Biochem. Biophys. Acta 470: 17-34.

Henry, D. R. and J. H. Block. 1979. Classification o drugs by discriminant analysis using fragment molecular connectivity values. J. Med. Chrm. 22: 465-472.

Hesser, C. M., L. Fagraeus, and J. Adolfson. 1978. Roles of nitrogen, oxygen and carbon dioxide in compressed air narcosis. Undersea Biomed. Res. 5: 391-400.

Heymans, F., L. Le Therizien, J.-J. Godfroid. 1979. Quantitative structure-activity relationships for N-[(N',N'-distributed-amino)acetyl] jarylamines for local anesthetic activity and acute toxicity. J. Med. Chem. 23: 184-195.

Hill, M. W. 1978. Interaction of lipid vesicles with anesthetics. Ann. N.Y. Acad., Sci. 308: 101-110.

Hopfinger, A. J., R. Potenzone, Jr., R. Pearlstein, O. Kikuchi, M. Shapiro, G. W. A. Milne, and B. R. Heller. 1979. Structure activity analyses int he classification of toxic chemicals. Proc. Sym. on Safe Handling of Chemial carcinogens. Ann Arbor Press, Ann Arbor, MI Vol 2, p 385-407, 1979.

Hudson, D. R., G. E. Bass and W. P. Purcell. 1970. Quantitative structure-activity modesl, some conditions for application and statistical interpretation. J. Med. Chrm. 13: 1184-1189.

Hunn, J. B. 1970. Dynamics of MS222 in the blood and brain of freshwater fishes during anesthesia. U.S. Bur. Sport Fish Wildlife 42: 1-8.

Hunn, J. B. and J. L. Allen. 1974. Movement of drugs across the gills of fishes. Ann. Rve. Pharmacol. 14: 47-55.

Iwamura, H. 1980. Structure-taste relationship of perillartine and nitro- and cyanoanline derivatives. J. Med. Chem. 23: 308-312.

Jain, M. K., J. Gleeson, A. Upreti, and G. C. Upreti. 1978. Intrinsic perturbing ability of alkanols in lipid bilayers. Biochem. Biophys. Acta 507: 1-8.

Jain, M. K. and L. V. Wray, Jr. 1978. Partition coefficients of alkanols in lipid bilayer/water. Biochem. Pharmacol. 27: 1294-1296.

Jain, M. K., N. Y. Wu. and L. V. Wray. 1975. Drug-induced phase change in bilayer as possible mode of action of membrane expanding drugs. Nature 255: 494-495.

Johnson, F. H., D. E. S. Brown, and D. A. Marsland. 1942. Pressure reversal of the actions of certain narcotics. J. Cell COmp. Physiol. 20: 269-276.

Jurs, P. C., B. R. Kowalski, and T. L. Isenbour. 1969. Computerized learning machines applied to chemical problems. Anal. CHem. 41: 21-27.

Kamaya, H., I. Ueda, and P. S. Moore. 1979. Antagonism between high pressure and anesthetics in the thermal phase-transition of dipalmitoyl phosphatidylcholine bilayer. Biochim. Biophys. Acta 550: 131-137.

Karickhoff, S. W., D. S. Brown and T. A. Scott. 1979. Sorption of hydrophobic pollutants on natural sediments. Water Res. 13: 241-248.

Kaufman, R. D. 1977. Biophysical mechanisms of anesthetic action: Historical perspective and review of current theories. Anesthesiology 46: 49-62.

Keynes, R. D. 1972. Excitable membranes. Nature 239: 29-32.

Kier, L. B., and L. H. Hall. 1976. Molecular Connectivity in Chemistry and Drug Research. Academic Press, Inc., New York, N.Y. 257 pp.

Klontz, G. W. and L. S. Smith. 1968. Methods of using fish as biological research subjects in method of animal experimentation (ed. Gay, W. I.) Vol. 3, pp. 323-379. Academic Press, New York.

Koblin, D. D. and E. I. Eger. 1979. Current concepts theories of narcosis. N. Engl. J. Med. 301(22): 1222-1224.

Kowalski, B. R. and C. F. Bender. 1974. The application of pattern recognition to screening prospective anticancer drugs. Adenocarcinoma 755 biological activity test. J. Amer. CHem. Soc. 93: 916-918.

Kubinyi, H. 1979. Lipophilicity and biological activity. Drug Res. 29(II), Nr. 8: 1067-1080.

Kubinyi, H. 1978. Non-linear dependence of biological activity on hydrophobic characteri: the bilinear model. Il Farmaco, Ed. Sc. 34: 248-276.

Kubinyi, H. 1977. Quantitative structure-activity relationships. 7. Tje bilinear model, a new model for nonlinear dependence of biological activity on hydrophobic character. J. Med. Chem. 20: 625-629.

Kubinyi, H. and O.-H. Kehrhahn. 1976a. Quantitative structure-activity relationships. 1. The modified Free-Wilson approach. J. Med. Chem. 19: 578-586.

Kubinyi, H. and O.-H. Kehrhahn. 1976b. Quantitative structure-activity relationships. 3. A comparison of different Free-Wilson models. J. Med. Chem. 19(8): 1042-1049.

Kuschinsky, G. and Lullmann, H. 1973. Textbook of Pharmacology, Chapter 4. Acad. Press Inc., New York, N. Y. 1-407.

Lee, A. G. 1976. Model for action of local anaesthetics. Nature 262: 545-548.

Lever, M. J., K. W. Miller, W. D. M. Paton, and E. B. Smith. 1971. Pressure reversal of anesthesia. Nature 231: 368-371.

Llinas, R. R. and E. J. Henser. 1977. Depolarization-release coupling systems in neurons. Neurosci. Res. Program Bull. 15: 556-687.

Longnecker, D. E. and P. D. Harris. 1980. Microcirculatory actions of general anesthetics. Fed. Proceed. 39: 1580-1583.

Malinowski, E. R. 1977a. Determination of the number of factors and the experimental error in a data matrix. Anal. Chem. 49(4): 612-617.

Malinowski, E. R. 1977b. Theory of error in factor analysis. Anal. chem. 49(4): 606-612.

Martin, Y. C. and H. N. Panas. 1979. Mathematical considerations in series design. J. Med. Chem. 22: 784-791.

Marvin, D. E. and D. T. Burton. 1973. Cardiac and respiratory responses of rainbow trout, blue gills and brown bullhead catfish during rapid hypoxia and recovery under normaxic conditions. Comp. Biochem. Physio. 46A: 755-766.

McCreery, M. J. and W. A. Hunt. 1978. Physico-chemical correlates of alcohol intoxication. Neuropharmacol. 17: 451-461.

McElroy, W. D. 1947. Mechanism of inhibition of cellular activity by narcotics. Quart. Rev. Biol. 22: 25-58.

McFarland, J. W. 1970. On the parabolic relationship between drug potency and hydrophobicity. J. Med. CHem. 13: 1192-1196.

McFarland, W. N. and G. W. Klontz. 1969. Anesthesia in fishes. Fed. Proceed. 28: 1535-1540.

Meyer, K. H. 1899. Zur Theorie der Alkoholnarkose. Arch. Exper. Path. U. Pharmakol. 42: 109-118.

Meyer, K. H. 1937. Contributions to the theory of narcosis. Trans. Faraday Soc. 33: 1062-1068.

Miller, K. W., W. D. M. Paton, and R. A. Smith. 1973. The pressure reversal of general anesthesia and the critical volume hypothesis. Mol. Pharmacol. 9: 131-143.

Miller, J. C. and K. W. Miller. 1975. Approaches to the mechanisms of action of general anesthetics. Blaschko, H. F. K. ed., Physiological and pharmacological biochemistry. Biochem, Series One, Vol. 12, Publisher Butterworth, london, pp. 33-76.

Miller, K. W. 1977. Towards the molecular bases of anesthetic action. Anesthesiology 46: 2-4.

Miller, K. W. and K. Y. Y. Pang. 1976. General anesthetics can selectively perturb lipid bilayer membranes. Nature 263: 253-255.

Morrison, D. G. 1969. On the interpretation of discriminant analysis. J. Market. Res. 6: 156-163.

Mullins, L. J. 1954. Some physical mechanisms in narcosis. Chem. Rev. 54: 289-323.

Ogino, A., S. Matsumura, and T. Fujita. 1980. Structure-activity study of antiulcerous and antiinflammatory drugs by discriminant analysis. J. Med. Chrm. 23: 437-444.

Operating Instructions Microtox™ Toxicity Analyzer Model 2055, Interim Manual Number 110679, Beckman Instruments, Inc., Microbics Operations, Carlsbad, CA, 1979.

Otto, P., M. Steel, J. Ladiki, R. Muller. 1979. From retrospective to predictive structure-activity correlations. J. Theor. Biol. 78: 197-210.

Overton, E. 1901. Studien uber de narcose Zugleich ein Beitrag zur allgemeinen pharmakologie Jena: Verlag von Gustav Fischer: 1901.

Pauling, L. 1961. Molecular theory of general anesthesia. Science 134: 15-21.

Price, H. L. and S. T. Ohnishi. 1980. Effects of anesthetics on heart. Fed. Proceed. 39: 1575-1579.

Pringle, M. J. and K. W. Miller. 1978. Structural isomers of tetradecenol discriminate between lipid fluidity and phase transition theories of anesthesia. Biochem. Biophys. Res. Commun. 85: 1192-1198.

Rahwan, R. G. 1975. Toxic effects of ethanol: possible role of acetaldehyde, tetrahydroisoquinolines and tetrahydro-B-carbolines. Toxicol. Appl. Pharmacol. 34: 3-27.

Rasmussen, G. T., T. L. Isenhour, S. R. Lowry, and G. L. Ritter. 1978. Principal component anlaysis of the infrared spectra of mixtures. Anal. Chim. Acta 103: 213-221.

Richards, C. D., K. Martin, S. Gregory, C. A. Keightley, T. R. Hesketh, G. A. Smith, G. B. Warren and J. C. Metcalfe. 1978. Degenerate perturbations of protein structure as the mechanism of anesthetic actions. Nature 276: 775-779.

Rimon, G., E. Hanski, S. Braun, and A. Levitzki. 1978. Mode of coupling between hormone receptors and adenylate cyclase elucidated by modulation of membrane fluidity. Nature 276: 394-396.

Rosenberg, Ph. H., E. Hansjorg, and S. Anton. 1975. Biphasic effects of halothane on phospholipid and synaptic plasma membranes: a spin label study. Mol. Pharmacol. 11: 879-888.

Roth, S. H. 1979. Physical mechanism of anesthesia. Ann. Rev. Pharmacol. Toxcol. 19: 159-178.

Roth, S. H. 1980. Membrane and cellular actions of anesthetic agents. Fed. Proceed. 39: 1595-1599.

Roth, S. H. and P. Seeman. 1971. All lipid soluble anesthetics protect red cells. Nature New Biol. 231: 284-285.

Roth, S. H. and P. Seeman. 1972a. Anesthetics expand erythrocyte membranes without causing loss of potassium. Biochem. Biophys. Acta 255: 190-198.

Roth, S. and P. Seeman. 1972b. The membrane concentrations of neutral and positive anesthetics (alcohols, chlorpromzaine, morphine) fit the Meyer-Overton rule of anesthesia: negative narcosis do not. Biochem. Biophys. Acta 255: 207-219.

Santara, N. J. and K. Augary. 1975. Non-computer approach to structure-activity study. An expanded fibonacci search applied to structurally diverse types of compounds. J. Med. Chem. 18: 959-963.

Saxberg, B. E.H., D. E. Duewer, J. L. Booker and B. R. Kowalski. 1978. Pattern recognition and blind assay techniques applied to forensic separation of whiskies. Anal. Chim. Acta 103: 201-212.

Schaad, L. J. and B. A. Hess, Jr. 1977. Theory of linear equations as applied to quantitative structure-activity correlations. J. Med. Chem. 20: 619-625.

Schettini, A. and W. W. Furniss. 1979. Brain water and electrolyte distribution during the inhalation of halothane. Br. J. Anaesth. 51: 1117-1124.

Schwab, S. H. and D. Schwab. 1979. The transformation of cytoplasmic microtubules into helices and paracrystals by halothane in the foraminifer Allogromi laticollaris. Arnold Z. Mikrosk Anat. Forsch 93(4): 751-762.

Seeman, P. 1972. Membrane actions of anesthetics and tranquillizers. Pharmacol. Rev. 24: 583-655.

Smith, F. M. and D. R. Jones. 1978. Localization of receptors causing hypoxic bradycardia in trout (Salmo gairdneri). Can. J. Zool. 56: 1260-1265.

Stenger, V. G. and T. H. Maren. 1974. The pharmacology of MS222 (ethyl m-aminobenzoate) in Squalus acanthias. Comp. Gen. Pharmacol. 5: 23-35.

Stockard J. and R. Bickford. 1975. The neurophysiology of anesthesia. Gorden E., ed. A basis and practice of neuroanesthesia. Amsterdam: Excepta-Med. pp. 3-46.

Topliss, J. G. 1977. A manual method for applying the Hansch approach to drug design. J. Med. Chem. 20(4): 463-469.

Topliss, J. G. and R. P. Edwards. 1979. Chance factors in studies of quantitative structure-activity relationships. J. Med. Chem. 22: 1238-1244.

Traube, J. 1904. Theorie der Osmose und Narkose. Arch. Geo. Physiol. 105: 541-558.

Trudell, J. R. 1977. A unitary theory of anesthesia based on lateral phase separations in the nerve membranes. Anesthesiology 46: 5-10.

Trudell, J. R., D. G. Payan, and J. H. Chin. 1975. The antagonistic effect of an inhalation anesthetic and high pressure on the phase diagram of mixed dipalmitoyl-dimyristoylphosphatidylcholine bilayers,. Proceed. Natl. Acad. Sci. USA 72: 201-210.

Unger, S. H. and C. Hansch. 1973. On model building in structure-activity relationships. A reexamination of adrenergic blocking activitr of β-Halo-β-arylalkylamines. J. Med. Chem. 16(7): 745-749.

U.S. EPA. 1978. Experimental design, application and data interpretation protocol. U.S. EPA, ORD-Corvallis, OR. EPA-600/9/78-018. 126 pp.

Vandam, L. D. 1965. Uptake and transport of anesthetics and stages of anesthesia IN: Drill's Pharmacology in Medicine (ed. J. R. DePalma), Blakiston Div. McGraw-Hill Book Co., New York, 3rd edition, pp. 85-99.

Veith, G. D., N. M. Austin, and R. T. Morris. 1979. A rapid method for estimating log P for organic chemicals. Water Res. 13: 43-47.

Veith, G. D., D. L. DeFoe, and B. V. Bergstedt. 1979. Measuring and estimating the bioconcentration factor of chemicals in fish. J. Fish. Res. Board Can. 36: 1040-1048.

Warren, G. B., M. D. Housley, J. C. Metcalfe, and N. J. M. Birdsall. 1975. Cholesterol is excluded from the phospholipid annulus surrounding an active calcium transport protein. Nature 255: 684-687.

Weisel, C. P. and J. L. Fasching. 1977. Deceptive "correct" separation by the linear learning machine. Anal. Chem. 49: 2114-2116.

Winters, W. D. 1976. Effects of drugs on the electrical activity of the brain: anesthetics. Ann. Rev. Pharmacol. 413-426.

Wu, R. S., M. K. Wolpert-DeFilippes, and F. R. Quinn. 1980. Quantitative structure-activity correlations of rifamycins as inhibitors of viral RNA-directed DNA polymerase and mammalian α and β DNA polymerases. J. Med. Chem. 23: 256-261.

Yalkowsky, S. H. and G. L. Flynn. 1973. Transport of alkyl homologs across synthetic and biological membranes: A new model for chain length-activity relationships. J. Pharm. Sci. 62(2): 210-217.

Zaleska, M. 1979. Effect of pentobarbital anaesthesia on serotonin metabolism in hypoxic rat brain. Pol. J. Pharmacol. Pharm. 31(3): 179-186.